入門經濟學叢書

統計學入門

寺尾琢磨著

廣文社

統計學入門

寺尾琢磨著

入門經濟學叢書

東京廣文社刊行

はしがき

經濟學も、それが一ケの學問である以上は、簡單化と抽象化の過程を必要とすることは勿論である。しかしそのため經濟學が現實と遊離した虛空の存在物となつてよい筈はない。經濟學はどこまでも經驗科學であり、現實と密接に結びついておらねばならない。戰時に流行した日本經濟學の如き理念的神話的經濟學が、結局何物をも齎らさなかつたことは餘りにも當然であつた。これを、ひたすら現實の分析に沈潛したアメリカのそれと較べれば、勝敗は最初から約束されてゐたといつてよい。

われわれは既に戰爭は放棄したが、さりとて經濟學がもとのまゝでよい筈はない。それは戰爭と切離されたゞけで、それだけ別の任務が大きく浮び上つて來たのである。壞滅した國民經濟を再建し、世界に公約した文化國家の物的基礎を獲得することは、われわれに課せられた貴いそして最大の義務であつて、これを可能ならしめるものは、現實に出發し現實を直視する經濟學を措いてはない。

では經濟學はいかにして現實と結びつきうるか。他なし、統計を通じてゞある。社會的な、即ち大きな集團的事實は、統計によらなければ正確に把捉できないからである。統計を必要とする

ものは素より經濟學に限らない。他の社會科學も、否、大部分の自然科學も亦同樣である。この意味に於て、統計は凡ゆる科學の要件であり、從つて統計の原理を說明する統計學は凡ゆる科學の共通財產といへるであらう。

本書は入門經濟學叢書の一部たる關係から、統計學を特に經濟學との關聯に於て出來るだけ平易に述べたものである。平易を主眼としたが、基本的な原理と方法は盡したつもりである。簡單な統計的硏究ならこの程度で足りると思はれるが、私の念願は、讀者が本書によつて統計學的考へ方の必要を覺り、統計學の本格的硏究に着手されることである。

私は慶應義塾で最も雜用の多い者の一人で、之に加へて最近新制高等學校長に任命されたゝめ、この數ケ月は文字通り寸暇なく、本書の校正も書肆に一任する外なかつた。故に原稿に於ける誤記を校正に際して訂正するチャンスを逸したわけで、少からぬ不安を感ぜざるを得ない。もし誤りがあれば、再版の際に訂正するつもりであるが、この點に關し讀者の御注意を得ることが出來れば、甚だ幸である。

昭和二十三年四月

三鷹臺の隅居にて

寺　尾　琢　磨

目次

統計學入門

第一章　緒　論

一　比較の重要性とその前提

とかく雨傘というものは荷厄介なものである。なしで濟めばこれに越したことはないが、それを敢へて持ち歩くのは、濡れるよりはましだからである。現に降つてゐれば問題はないが、たゞ空模樣が怪しいといふ程度だと、持つてゆかうかゆくまいか、一と思案というところ。思案とは持たずに行つて濡れる不快さと持つて行つて無駄になる不快さとを心の中で比較することである。思案は個人だけの問題ではない。運輸省としては運賃は十倍も値上げしたいらしいが、それによる物價騰貴や大衆の不滿を考へれば、そう無駄な値上げもできないであらう。何倍にしたらよいか、これは極めて複雜な比較によつてのみ始めて決定されることである。

何れにしろ個人の場合でも政府の場合でも、或る事柄を決定するには、必ずそれによつて得る

ものと失ふものとを比較して正しい判斷を下さねばならぬ。これをしない行動は凡べて行當りばつたりの衝動的な、卽ち本能的かまたは狂人的な行動で、合理的行動ではない。そして人類の偉大な進歩が一に人間の合理的行動の結果だとすれば、この比較といふ人間の腦裏に展開される取捨選擇の闘爭こそ、文化とその發展の源泉と言はねばならぬ。

勿論日常生活で絶えず出會ふ簡單な事柄については、比較はいつしか習慣的に卽ち半ば本能的に行はれ、特にこれを行つてゐるといふ意識はない。雨が降つてゐれば躊躇なく傘をもつてゆく。驛までは無意識に同じ道を歩いてゆく。ほかの道の方がいゝかどうかは考へない。併し多少とも重要な事柄については、それが日常のことであつても、多かれ少かれ判斷を要する。卽ち比較を要する。ものを買ふには必ず對價を支拂はねばならないから、唯しも一應買ふか買ふまいかを反省するであらう。ことが重大になればなるほど、この反省は大きくなる。家を建てようか建てまいか、普通の人なら隨分頭を惱ますに相違ない。彼が家を建てたら、これは、彼が家を建てる方が得だと判斷した證據で、かように人の行爲の結果は判斷を反映してゐるのである。かくてわれわれは結果から遡つて、原因たる判斷卽ち比較の狀態を推しうるのであつて、パレートの選擇理論はこの原理によつて效用可測性の問題を解決したものである。それは兎に角として、われわ

れの合理的行動が正しい比較の上に立つてゐるといふことには渝りがない。これによつて正しい判断と、延いて正しい行動が可能となるからである。では正しい比較はいかにして可能であるか。

原理的にはその答は簡單である。比較される双方を正確に知るといふことである。傘の厄介さと濡れる不快さを、家を建てる犠牲と家をもつ利益を、はつきりと捉へないでどうして決斷が下せよう。その一方だけ知つて他方を知らなければ、たとへ一應の判斷は下せるにしても、決して正しい判斷とはならず、從つて正しい行動の指針にはならない。傘の厄介さだけ知つて濡れる不快さを知らなければ、雨の降るたびに濡れてゐなければならない。家をもつ愉快さだけ知つて建てる犠牲を知らなければ、借金で首がまわらなくならう。不利だけ知つて利を知らない人は臆病になり、利だけ知つて不利を知らない人は無鐵砲になる。蓄めることを知つて出すことを知らなければ吝嗇になり、出すことは知つて蓄めることを知らなければ浪費家になる。正しい行動の指針としての比較が、關係事項に關する正確な知識を前提とすることは、最早や、贅言を要しまい。

そこで問題は、では正確とは何か、そしてどうすればそれを得られるかに歸着する。端的にい

へば、正確とは數字的といふことであり、從つて、正確に知るとは、對象を數字で表現するといふことである。鎌倉の大佛は大きいとか、木綿糸は弱い、と言つた表現は、必ずしも嘘ではない。併し大小强弱といふ言葉は甚だ主觀的で、人によつてどのようにも解される。鎌倉の大佛より大きいもの、木綿糸よりも弱いものは、探せばいくらでもある。奈良の大佛と較べれば鎌倉の大佛は小さく、スフ糸と較べれば木綿糸は强いと言はねばならぬ。比較を超越した絶對物例へば眞理とか神とかは別として、物には必ず程度がある。程度を明瞭に表現することが正確といふことなのである。それには單に大小强弱といつた主觀的言葉では足りないのであつて、一定の客觀的標準が必要である。かゝる標準は計數または計量によつてのみ與へられる。一定の單位に從つて數へまたは測つた結果は必ず數字として示される。それによつて一物の大小强弱等々の程度は一義的に決定される。（計數と計量の相違については後に述べる。）數字は事物を正確に表現する唯一の手段なのである。逆にいへば、計數または計量の不可能な、即ち數字的に表現され得ない事柄は多かれ少かれ不正確である。善惡美醜といつた人の感情なぞ、いつの世になつても數字的に表現することはできないであろう。百%の美人とか、五十%の惡人といつた文句は、一つの修辭の外の何物でもない。

二　數字化の方法とその困難さ

そこで正確な知識は數字的知識に外ならぬことが明かだとして、問題はいかにして對象を數字化しうるかといふことである。勿論それが計ることによつてのみ可能なことは言ふ迄もないが、計る方法は對象の性質によつて多種多樣である。先づ第一に、計數と計量の別がある。一つ二つと數へるのは計數で、物指・衡り・桝などで計るのは計量である。計數は物件の數を、計量はその量を計る手段と言へる。これは謂はゞ自明の理であるが、動もすれば閑却され易い二三の點がある。その一つは、元來數へうるものでも、單位が餘り小さいときは、寧ろ量として計るといふことで、例へば米粒は一つ二つと數へられるが、實際には桝や目方で、即ち量として計る。逆に量もそれが一定單位の大さに纒められゝば計數できるのであつて、ビン詰の酒は一本二本と數へられる。これらは總べて實生活の便宜からきたことで、數と量との區別はそれほどはつきりしたものでないことが判る。次の注意は、計數は單位の數を言へばよいから、甚だはつきりしてゐるが、計量は計器の目盛を讀むことで、眞の意味の正確さは得られないといふことである。五人は誰が數へても五人だが、五尺は果して正確に五尺かどうか、多分に疑問がある。物指自身、多少

の狂ひがあらう。濕度による多少の延び縮みさへ馬鹿にできない。また、たとへ計器は正確でも、全く正確に目盛を讀むといふことは、決して出來るものではない。目盛は線である。線は理論的に長さのみあつて幅はない筈だが、それでは人目に映らないから、人に示す爲には或る幅をもたせねばならぬ。卽ち目盛そのものが既に正確を缺いてゐるが、加之、目盛と目盛との間は空白で、結局われわれは目測で大凡の見當をつける外はない。勿論計器は無限に精密となり、一粍の何千分の一といふ小さいものも計れるが、それすら猶ほ微小の誤差は免れず、而も一般にはそんな精密器は使用されてゐない。要するに目方と長さと容量の三者は多かれ少かれ不正確で、所謂近似値に過ぎないのである。勿論實生活ではそれほど不便は與へないかも知れない。知れないどころか、實はそのために生活が順調に運んでゆくのである。正確さに餘りに捉はれては、一刻たりとも安閑としてはゐられない。百匁の牛肉が一匁足りないといつて一々目の色をかへねばならないとしたら、どうであらう。正確を尊重するのはよいが、さりとて正確狂になつてはお終ひである。唯だ併し、百匹の牛といふものはあるが、百匁の牛肉といふものが決してあり得ないといふことは、忘れてならぬことである。

ないことは言ふまでもない。特別の不便が起らない限り、何事も正確なればなるほどよいといふこと、そしてこれは數字化によつてのみ期待されるといふことは、依然眞實である。そして對象の數字化は計數又は計量によつて兎に角可能である。但しそのためには對象が物質であること丶、手近にあることを必要とする。ところが世の中にはこの要件を具へてゐないものは甚だ多い。これらについて測定を斷念せねばならぬとすれば、われわれのもち得る正確な知識といふもの丶範圍は甚だ限定されるであらう。卽ち問題は第一には質、第二には集團的事實に關する場合である。

質は元來數量に對する言葉であつて、計測できないのが原則である。併し質にも必ず程度がある以上、これを何等かの形で現はす必要がある。一般にはこれを種々な違つた言葉で言ひ現はす。美しさの程度を、最も大雜把には善と惡の二つで片づけ、それでは餘り大雜把すぎる場合には、これに「最も」「甚だ」「可成り」「少し」「ほんの少し」等々の副詞をつけるとか、優、良、可、不可、のような階段的評語を用ひる。この考へ方を撤底させて、最高を百、最低を零とし、その間の數字によつて區別することもできる。答案の可否を八〇點とか三五點とかの評點で示すのはその例である。かような例は枚擧に遑がないのであつて、殆ど凡ゆる質の程度は、かような手段によつて數字で示すことができよう。併しこの場合の數字は、計數または計量といつた

機械的方法から生れた客觀的な數字とは根本的に性質を異にし、要するに前述の副詞的敍述と本質的には相違のない主觀的表現に過ぎない。そこには數字的表現の基本條件たる客觀的標準が缺けてゐるのである。而も、それが數字的外見をもつため、恰も何等かの客觀的標準が存在するような錯覺を起させ、不當に信用させる場合が多いのである。數字のかような魔術性に對しては嚴重警戒せねばならぬが、同時にこの種の數字化に於て出來るだけ客觀性をもたせる工夫が必要である。その手段としては、第一に質的現象にもできるだけ科學的標準を設けることである。昔はたゞ「明るい」とか「暗い」とかしか言へなかつた明暗の程度も今ではルックスといふ光度單位で計ることができる。適性檢査がもつと發達すれば、人の技能も客觀的に規定されることにならう。このことから判ることは、質と量とは必ずしも始めから嚴重に區別されてゐるものではなく、人の智慧が進むに從つて、質は次第に量に轉化するといふことである。このことは人の經濟行爲に於て最も明瞭に現はれてゐる。經濟とは元來欲望滿足といふ質的事實であるが、貨幣經濟の下では總べては價格といふ量的事實に轉化されて了つた。よい品は高く、惡い品は安い。質の善惡は價格の上に示される。それが極端に應用されゝば、例へば子供や貞操までが價格で評價されれ取引されることになる。その可否は別として、兎に角貨幣といふ客觀的基準が支配的威力を振

つてゐる經濟現象は、少くとも外見上は量の世界に所屬するのである。

第二に必要なことは、たとへ客觀的標準が設けられても、もし光度の測定のように機械的に計れない場合には、同一對象を多くの人に評價させて平均的評價を求めることである。それによつて不當な主觀的判斷の誤謬を多かれ少かれ訂正できるからで、ミス東京の選出や水泳飛込競技の採點は大勢の試驗官が行ふのが恒例である。ではわれわれにとつて最も問題となる價格といふ量的表現は果して客觀性をもつてゐるであらうか。普通の商品の價格は、社會的に、即ち多數人の取引を通じて、決定されるから、謂はゞ極めて大勢の試驗官の採點から決定されたようなもので、著しい客觀性をもつてゐると言はねばならぬ。このことは闇價格にもいつの間にか相場ができて了ふことから判らう。勞働價格即ち賃銀や其他大部分の價格はこれに屬する。ところが自由競爭の行はれないものや、前述の貞操のように滅多に取引されないものについては、素より客觀性は乏しい。何れにしろ價格といふ數字を、本來の意味に於ける數字と全く同一視すべきでないことは、經濟學徒の常に銘記せねばならぬことである。

次の問題は、集團的事實の數字的表現といふことである。一つ又は少數の對象なら容易に數字化できるものでも、それが大きな集團となると種々の故障にぶつかるものである。これは結局人

（認識力の及ぶ範圍が案外狹いからで、簡單に一目で見渡せないものについては、大雜把な推定さへできない場合が少くない。東京都にどれだけの人口があるか、大阪市にどれだけの失業者がゐるか、物價はどれほど騰貴したか、これらの間に對しては普通の方法では答へられない。而も經濟學其他の社會科學に現はれる概念は殆ど常にかような集團的性質のものであるから、もしそれらが數字化されないとすれば、社會科學は殆ど常に「正確」といふ要請からは絶緣された抽象的學問に留まらねばならない。事實比較的最近まで經濟學は對象の數字化を閑却し、演繹的方法卽ち普遍的公理から論理的に結論を導き出す方法に主力を注いできた。正統學派や數理學派が「經濟人」の假定から自由競爭の原理や限界效用均等の法則を導いたり、需要量を以て價格の減少函數 $D＝\phi(p)$ と規定したことは、諸君の知る通りである。併しかような方法でどの程度まで經濟的事實を記述し説明しうるであらうか。需要量が價格と反對方向に動く傾きをもつことは事實としても、その情況卽ち需要彈力性は商品により、時により、處によつて相違する。單に $D＝\phi(p)$ といつた函數の一般形では、又は單に dD/dp と記した需要彈力性では、この間の消息は全く不明である。それが具體的にいかなる形であるか、これが問題なのであつて、そのためには實際の價格と實際の需要量卽ち數字的資料から、數字的にこの函數形を具體化せねばならぬ。

米價が安くなれば需要は增加するといふ一般論なら、抽象的需要法則から容易に導かれるが、一割安くなつたら需要量は何割增加するかといふ問題は、具體的數字的資料を以てしなければ答へられないことである。

これらの問題に於ける數量は一商店の價格、一家庭の需要量ではなくて、何れも社會的集團的數量である。かゝる數量を數字的に表現したものを特に統計といふ。統計は數字であるが、一般の數字とちがふ點は、集團的事實を示すための數字だといふことである。「統計は數字であるが、總べての數字必ずしも統計でない」といふことをはつきり掴んでおくことが必要である。では統計の主體たる集團的事實とは何か、またそれを數字化するには如何なる手段をとらねばならぬか。われわれはこれを問題とすることによつて、次第に統計學の本質に接近してゆくわけである。

第二章　統計調査

一　統計集團

數量的事實といつても最初から數字として與へられてゐるわけではなく、われわれが之を數へ又は量ることによつて始めて數字となるのである。この過程は、箇別的事實については比較的簡單であるが、集團的事實については複雜な手續が必要となる。人口を數へるのに、テーブルの上の蜜柑のように一つ二つと數へることはできない。豫め調査用紙を配布し、一定時刻の現狀を記入せしめ、これを取纒めて所謂集計せねばならぬ。集團を數字化するには、即ち統計を作るには、かような手續が必要で、これを統計調査といふ。それは一般に多くの準備・費用・人員を要し、ときには法律的强制力をも必要とするから、主として官廳の仕事である。併し小さな集團については私人的調査も不可能ではない。殊に最近では勞働組合が所屬員の生計調査を行つたり、學校が學生の實態調査を行つて、生きた參考資料を得ようとする傾きが盛んとなつた。統計は必ずしも常に上から與へられるものとは限らないのである。英米流の統計學は統計解析を對象と

し、調査に關する說明は殆ど或は全く省略されてゐるが、資料たる統計そのものがいかにして獲得されたかは、資料そのものゝ性質に大きな影響を與へるから、解析の前程としてもその性質を辨へておく必要がある。

統計とは集團的事實の數字的表現である。統計によつて表現されうる集團的事實を統計集團と名づける。統計集團とは、「概念的に同種で、時間的にまた場所的に限定された多數の單位の全體」を意味する。世の中に凡ゆる點で全く等しいものはないが、われわれは便宜上、或觀點から同一と認められるものを一括して特定の種類をつくり上げる。君と僕とは總べての點で必ずしも等しくはないが、人間といふ動物で且つ日本に國籍をもつてゐるといふ點では同じである。そこで人間で且つ日本に國籍をもつところのものを一括して日本人と名づける。品物は無數にあるが、それらのうち賣買されるものを一括して商品又は財と名づける。かように一括して特定の名稱を與へることを一般に概念を構成するといふ。概念は抽象的なこと、例へば善とか美とかいつたことについても構成されるし、また時間的場所的制限を必要としない。ところが對象を具體物に且つ時間的場所的に限定すれば、概念は計數または計量の可能な大さとなる。これを統計集團といふのである。こゝで注意しなければならぬことは、一括する場合の觀點は種々樣

々で、従つて二つのものは時には同種となり時には異種となるといふことである。人間といふことと國籍といふこととの外に更に性といふことを考へれば、日本男子と日本女子といふ二つの種類ができ上る。賣買されるといふことの外に、用途を考へれば、例へば生産用商品（生産財）と消費用商品（消費財）が區別される。そしてこれらは更にいくつもの種類に分つことができるのであつて、統計集團はこれこれのものでなければならないといふ規定は全くないのである。卽ち一つの統計集團はより大きい統計集團の一部分たりうると共に、より小さい統計集團の集りとも見られるのである。某金屬工場の勞働者は、それだけで一つの統計集團となるが、更に細分してその工場の男工と女工といふ二つの統計集團とすることもでき、また同種の工場全部を合せた日本金屬工業勞働者といふ大きな統計集團の一部ともなり、或ひは日本工業勞働者といふより大きい統計集團の一部ともなりうる。

概念的に同種と認められたものでも、時間的にまた場所的に限定されなければ統計集團にはならない。漫然たる集團は集合體ではあるが、數字的に表現する術はないから、統計集團ではない。例へば唯だ日本人といつただけでは摑みようがない。何千年も前から日本に生れ日本に死んだ人は日本人といふ種類に入るが、それが幾人になるか、全く調べようもない。また、時を現在

に限定しても、現に外國にゐる日本人については調査は殆ど不可能である。斯くて日本人を統計集團とする場合には、日本といふ領土に特定期日に又は特定期間存在する日本人に限定する外はない。併しこの時間的場所的制限も亦、豫め一定の規則があるわけでなく、必要に應じまた對象の性質によつて定められるのである。出生や死亡の如き現象は瞬間的事件であるから、これを記錄するには一定期間（週・月・年等）內の生起回數をとらねばならぬ。災害、罷業、輸出入額等も亦同じである。

即ち週別・月別・年別等の所謂動態統計が得られる。之に反して人口の如き持續的で不斷に變化してゐるものについては、特定瞬間に於ける狀態を求める外はない。これを靜態統計といふ。靜態統計は特別の統計調査を必要とし、國勢調査や工場調査はその例である。然るに動態統計には殆ど常に、他の目的のために作られた記錄を利用する。例へば出生は出生の都度屆出でられるが、それは行政的必要から行はれることで、特に出生統計を作るためではない。併しそれを集計すれば出生統計が得られるのであつて、恰も最初から出生統計を作るために出生屆を命じてあると同じ結果になる。かように他の目的に集められた資料を利用して作つた統計を二次統計といひ、最初から統計作製を目的とした特別の調査を行つて得られる一次統計と區別する。二次統計

の場合にも、何を以て統計集團とするかは明瞭に確定されねばならない。

二　調査の種類

統計集團は規定の如何によつて大きくも小さくもなることは前述したが、一度び決定された統計集團は、これを構成する全單位について調査することによつて完全な統計となる。これを悉皆調査といふ。これが本來の意味の統計調査であつて、國勢調査はその適例である。ところが全單位の代りに一部を調査し、その結果から全體を結論する場合がある。これを部分調査といふ。部分調査では集團の大さは知り得ないのが原則である。東京都の人口數だけから全國人口を知ることはできないであらう。尤も何かの理由で、その部分が全體に對して占める割合が判つてゐれば別であるが、かゝる割合は結局は統計調査から求める外はないから、一度も統計調査の行はれたことのない事柄については、部分調査から全體の大きさを知り得ないといふことは一般に言へることである。

併し統計的研究に於て、全體の大さは格別問題にならぬ場合が多い。賃銀統計の中心問題は賃銀が一般にどの程度で、またどの程度に變化したかであつて、これには單位たる賃銀勞働者の總

數が判つてゐなくても、方法如何では可成り正確に判ることである。一般に悉皆調査は費用・勞力等々の點で不可能な場合が多く、また迅速に結果を知るためには、調査範圍は狹い方がよい。勿論部分から全體を推すときは、多かれ少かれ誤差の生ずるを免れないが、最近の統計學の進歩は、これから起る困難を合理的に解決しうる途を開くに至つた。今後の統計調査が部分調査の形で發達するであらうことは確實である。

部分調査はいくらかの種類に分たれる。一般に統計學では部分調査で扱ふ部分を試料といひ、これから推定しようとする全體を母域または母集團といふ。故に問題はいかにして試料を決定するかといふことである。詳細にこれに立入ることはこゝでは許されないが、主たるものは抽出調査と計畫的試料調査である。抽出調査とは公平に、即ち調査員の主觀によつて左右されないように、一部分を選出することである。液體を豫めよく攪拌して一部をとれば、その一部は全體のもつ成分を同じ割合で抱含する。故に試驗管へとつた小量を檢査することによつて、よく全體の例へば一つの井戸の中の水全部の成分を知ることができる。一萬人の勞働者から同樣にして千人をとれば、この千人は一萬人を十分の一に縮小したものと見られ、從つて逆にその結果を一萬人に及ぼすことができよう。この場合、勞働者は液體ではないから、一萬人を直接かき混ぜることは

できない。併し例へば富籤の抽籤方法のやうなやり方を利用すれば、結果は同様である。公平にといふ意味は、一萬人から千人をとる場合、全員の誰もが「選ばれる確率は十分の一、從つて選ばれない確率は十分の九」といふことである。この場合に起る誤差は、後に述べる誤差法則によつて數字的に處理することができる。誤差法則は確率論の中心命題であるから、抽出調査は畢竟確率論の上にのみ可能である。

次に計畫的試料調査とは、例へば勞働者生計狀態を調査する場合、全勞働者を賃銀、業種等によつていくつかの階級に分類し、各階級について部分調査を行ひ、最後にそれらを綜合する方法である。この方法を、英語では stratified sampling といふが、直譯すれば層化試料調査である。抽出調査では部分は可成り大きくなければならぬが、この方法によれば比較的少數の調査で足りる。但し階級の分け方や、各階級からとる試料の大さなぞによつて、結果は甚だしい影響を蒙るから、やり方いかんでは不公平な數字が生れる惧がある。Stratified Sampling が Strategical Sampling（戰略的試料調査）となつては困る。

右に述べた二つの方法が部分調査法の典型であるが、ときには標本調査と稱せられる極めて小部分についての方法がとられることがある。家計調査は小額所得者（勞働者や俸給生活者）の生

計狀態を知るために行はれるか、その際、數百乃至數千の標準家族を選定して、數ケ月乃至一ケ年に亙つて詳細に家計簿を記入せしめ、これを綜合して收支の一般狀態を算出する。家計狀態は家族員數によつて甚だ異るから、最も正常的な員數から成る家庭を選ぶのが最も效果的である。獨身者や多子家庭の家計は一般の標準とはなり得ないからで、一般に夫婦と子供二人乃至四人の家庭が選ばれるのである。かような標準家族は典型的單位であり、從つて試料といふよりは寧ろ標本である。標本調査と稱せられる所以であるが、實はかゝる調査は別に行はれたより大規模な調査を背景としてゐるのである。いかなる員數が正常的であるかは、先驗的には判らないことで、結局國勢調査その他の人口調査から求める外はない。昭和五年國勢調査の結果を見るに、その「人員別普通世帶及人口」は次の如くである。

世帶員	世帶	人員
一人	694,063	694,063
二人	1,480,773	2,961,546
三人	1,870,115	5,610,345
四人	1,905,489	7,621,956
五人	1,826,367	9,131,835
六人	1,596,536	9,579,216
七人	1,243,343	8,703,401
八人	851,617	6,812,936
九人	516,311	4,646 799
十人	297,722	2,977,220
十一人以上	317,940	4,021,504
計	12,600,276	62,760,821

即ち世帶數では四人世帶が、人員では六人世帶が最も多い。そこで四人乃至六人の家族が最も典型的なことが判るのである。かような資料が缺けてゐる場合は、標本調査は不可能である。但し家計調査は之を强制することができず、有志の應募に俟つ外はないから、この點で必ずしも標本調査にならぬ場合がある。綿密な家計簿の記入は概して生活の合理化された家庭にのみ行はれるから、自ら進んで應募するような家庭は特に合理化されたものが多く、手本にはなつても標本にはならないかも知れない。故にこの種の標本調査は寧ろ「不完全な部分調査」と名づけた方がよいとの說もある。併し兩者の區別は多くの場合甚だ困難である。

なほ部分調査の一種にアンケート(Enquête)と稱されるものがある。少數の關係者に特定事項の可否、贊否その他を問合せることで、輿論調査(世論調査)も亦多くの場合これに屬する。これが果して統計調査といへるかどうかは極めて疑問である。蓋し調査者の希望する答の得られそうな方面を選定することによつて、歪められた統計ができ上るからで、例へば勞働者の經營參加の可否を資本家に問合はせれば、恐らく大部分が否定的解答を寄せるであらう。解答者の屬する階層を明かにしないで、結果だけを發表すれば、世間一般がかゝる意見をもつかの如き印象を與へるであらう。言論の自由に乘じて「輿論」を振廻す傾向が甚だ强くなつたが、それが果して「

僞造された輿論」でないかどうか、大いに警戒を要することである。

併し近頃朝日新聞等で屢々行つてゐる世論調査は、相當綿密な層化試料調査法によつて、被調査者を年齡、體性、職業等の點で一方に偏しないよう努めてゐる。調査方法は發表の都度くはしく說明されてゐるから、こゝでは述べない。かゝる調査は舊來の單純なアンケートを遙かに越えたもので、本來の意味に於ける統計調査の一つに數ふべきであらう。

三　調査の障碍

統計集團を正確に記述するのが統計の目的だのに、動もすれば統計が噓の代名詞と考へられるのは、主として調査上の欠陷に由來する。單獨の對象を計ると違つて、大きな集團を計る場合に種々の困難が伴ふことは避け難いが、この困難は極力克服されねばならぬ。二三の要點を列記して見よう。

(一)概念は統一されねばならぬ。いま賃銀調査を行ふとし、第一に必要なことは、その場合の賃銀とは何かを確定することである。賃銀は抽象的には勞働に對する報酬と定義されたが、實際にはその內容は複雜多岐である。契約賃銀の外に種々様々な手當(殘業手當、家族手當)や實物給與

とか交通費又は税金の會社負擔といつた添物があるかと思へば、他方では組合費、保險金、税金、積立金等々が控除されるといつた狀態で、それが場所により時によつて違ふから、一口に賃銀といつただけでは判斷のしようがない。特に實物給與に至つては、正確な算定は殆ど不可能にちかい。故に調査に當つては、これら諸項目をいかに取入れるかを豫め確定しておかなければならない。同樣にして、失業調査では失業とは何かを、結婚調査では結婚とは何かを、先づ定めねばならない。失業や結婚に種々な形態のあることは誰しも知るところで、これを一般的抽象的規定で濟ませることはできない。婚姻動態統計は婚姻屆から作られるから、法律上の結婚だけが現はれ、靜態統計は事實上の配偶の有無に從ふから、所謂内緣關係も含まれる。有配偶者は男女同數で然るべきだが、一般に女の方が多い。離緣された女や妾が外聞を憚つて有配偶と申告し易いからである。

(二)人の名譽や利害に關することは把捉困難である。いま述べた配偶申告の例で判る通り、人は不名譽の事柄は正直に申告したがらない。前科、遺傳、性的事項等は聞く方が無理である。米國では、産兒制限を實行してゐるかどうか、又はいつ童貞を失つたかといつた事項を盛んに調査するが、その場合には理解ある人々のみを選擇せざるを得ないから、たとへ申告は得られても、一般

の事實は必ずしも反映されないであらう。
同様にして所得や利益は、所謂課税恐怖から概して内輪にしか申告されない。税務官は心得たもので必ず何割かを加算してくれる。反對に配給の基本を知るために人口調査を行ふと、必ず餘計な人口が申告される。所謂幽靈人口これである。農村では供出を誤間かすために耕地面積や收穫量を極力少く報告したがる。日本で實際にどれだけ米がとれるかは結局判らないといふのが實情かも知れない。配給、供出、課税といつた制度が存する限り、統計の正確が阻害されることは止むを得ない。興味あるのは女子年齢調査である。丁度區切りのよい年齢例へば二十歳と申告する者が甚だ多いが、これは二十プラスX歳の者が端數のX歳を切りすてるからで、若く思はれたい一念の發露であらう。逆に、非常な高齢者になると歳を誇張する傾きがある。百何歳といつた年齢には眉唾ものが多い。日本で年齢調査に生年月日を記入させるのは、一つには「數へ年」といふ不都合な數へ方がある爲であるが、主なる理由は、これによつて申告の嘘僞を防止するに在る。わざ〳〵計算してまで生年月日を誤間かそうとする人は餘りあるまい。
(三)餘り六ケ敷い又は多くの項目を調査してはならぬ。申告は當人の迷惑にはなつても、直接の利益は齎らさないのが普通であるから、簡單に濟ませたがるのは人情の常である。申告のために多

大の勞力や時間を必要とするような困難な問題や多くの項目は尋ねないのが利口であらう。調査側としてば「どうせついでだから」といつた氣持ちでつい餘計なことまで調べたくなるらしいが、國民の全部や大多數についての調査に於ては、特にこの原則が守られねばならない。但し敎育が普及し文化が進步するにつれて、この困難は次第に克服される。最近我國でも統計敎育の擴大を計畫してゐるとのことであるから、將來は國勢調査に際して多くの必要事項を調査することが出來ると思ふ。

第三章　統計表とグラフ

一　分類と統計系列

一つの統計調査が行はれたとする。例へば國勢調査なら、全國の千數百萬の世帶へ一枚づつ調査票を配布し、世帶全員の姓名・體性・年齢・配偶の有無・職業等々の調査事項（標識）を記入せしめこれを中央に回收する。そこには千數百萬枚の調査票が積み重ねられるわけで、これを整理し、分類し、製表して、始めて統計となるのである。二次統計なら調査は省略されるが、出生屆や賃銀支拂簿を集計する點では同じである。大規模な調査は巨大複雜な計算機にかけて機械的に處理する外はないが、工場や學校での調査なら、それほどの設備は要らない。一般に、いかなる形の統計ができ上るかは、實は調査の前に既に定められてゐるのである。卽ちいかなる表を作るかを豫定して調査事項がきめられるのであつて、詳細な統計を作らうとすれば質問そのものが詳細でなければならず、大雜把な統計でよいなら、質問そのものがこれに應じた粗雜なのが原則である。

整理の中心は分類である。分類は場所・時間・屬性の三つを標準として行はれる。靜態調査では時は一定してゐるから、場所と屬性だけが標準となる。國勢調査の結果たる全人口を都道府縣に分ち、更にその各々を郡市町村等に分つのは場所的分類で、かくしてできた一聯の數字を場所的統計系列といふ。また全人口を男女や年齢に分つのは屬性的分類であるが、この場合、屬性が事物的であるか數量的であるかに從つて事物的統計系列と度數的統計系列(度數系列)の差別が生ずる。男女別、職業別、配偶關係別などは前者の例で、年齢別は後者の例である。出生や死亡に關する動態調査は連續する時の間隔例へば月や年を標準とするから、月別や年別の、即ち時の經過に從つて分類された統計が生れる。これを時間的統計系列(時系列)といふ。

これらの分類に於て問題の多いのは度數のそれである。年齢や賃銀をどう刻んだらよいか、そこには一定の約束はないから、必要に應じてきめる外はない。そこで第一に間隔をどうするかゞ問題になり、第二にはその現はし方が問題になる。間隔は餘り細くすれば表が長くなり過ぎて見にくゝなり、餘り大きくすれば分類の效果が現はれない。年齢の場合には各歳別、五歳別、十歳別の外に、例へば十五歳以下と十五乃至六十歳と六十歳以上といつた分け方もある。これは全人口を幼少年・生産年齢者・老人に分けることで、便利な場合もあるが、一般にはこれでは大雜把

すぎる。そこで特別の目的が定められてゐないときは、寧ろなるべく細く分けておいた方がよい。必要とあればいくつかの間隔を一つに集めることもできるからで、始めから大きな間隔にしておくと、後にこれを分けることはできない。間隔の大さは階級または單に級と稱せられる。間隔は等しく刻むのが原則であるが、對象のいかんでは必ずしもこれに依る必要はない。例へば年齢別死亡率の表では屢々零歳階級だけ月別にしてある。これは生れた最初の一年は單に死亡が多いばかりか、月によつて大差があるからで、ときには最初の月は更に週別にさへしてある。また所得統計では高額になると次第に間隔を大きくする。でないと無闇に表が長くなるからである。

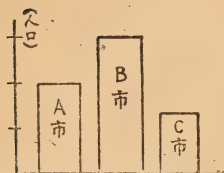
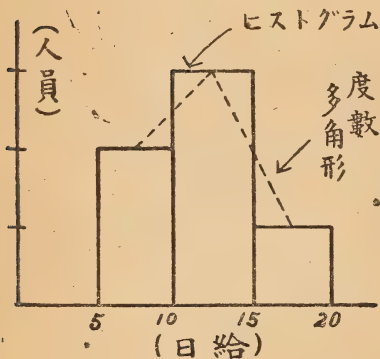
級の示し方は仲々六ケ敷い。年齢を 0—10, 10—20 と記せば、丁度十歳の者はどちらに屬するか判らない。故に 0—9,10—19 とした方がよい。金額の場合にも境をはつきりさせるためには、例へば 0—9.99, 10—19.99 とする。併し前後の關係から疑問の餘地がなければ、もつと簡單にしてもよい。

二 グ ラ フ

統計表は數字の羅列であるが、一般に數字を讀みとることは容易でない。著書や論文の中の統計表は敬遠され勝ちで、殊に資料展覽會などでは、統計表の前は素通りして了ふ人が多い。そこで數字を讀み易い形に變形する必要が起るのであつて、統計グラフはこの目的に應ぜんとするものである。

數字を長さや面積や體積で示すのは最も普通の方法である。數字を目盛に應じた長さの棒で示す棒圖表は甚だ無難であるが、圓や方形による面積圖表になると錯覺を起させ易い。圓の面積は半徑rの平方と圓周率πとの相乘積卽ち πr^2 であるから、二つの圓の大さは半徑の平方根に比例する。故に大きい數字ほど割合に小さな圓に見える。四萬圓と九萬圓は $\sqrt{4} : \sqrt{9}$ 卽ち2對3の半徑をもつ二つの圓で示される。正方形を用ひた場合も同樣に一邊の長さをそれぞれ2對3とせねばならぬ。餘程慣れた人でないと正確な比較は不可能である。面積の代りに體積例へば賽形の立方形を用ふれば、四對九の大さは一邊の長さが $\sqrt[3]{4} : \sqrt[3]{9}$ 卽ち1·587 : 2.080 の比となるよう描かねばならぬ。描くことが面倒なばかりか、見る人に正確な印象を與へることは益々望めなくなる。異る米の産額を大小の米俵で示す式の所謂繪畫圖表は複雜な體積圖表で、判讀は一層困難である。この場合には寧ろ同じ大さの米俵を例へば一方は五つ、他方は二つ描いて、五萬

石と二爲石を現はした方がよい。なほ圖表は元々人の目を惹きつけるのが目的であるから、適當な色彩を使へば一層效果的である。但し、色は錯覺を起させ易いから、その使ひ方は注意を要する。黑は白より小さく見え、電燈の下では薄黄色は白と區別がつきにくいものである。

場所的及び事物的系列は相互に獨立した棒を竝べることによつて圖示できるが（第一圖）、度數系列と時系列は、棒を用ひたときには間に空白のないようにし（第二圖）、一般には棒の頂點を連ねた線によつて示すことになつてゐる。分類標準たる級や時は連續した數字だからである。

度數棒圖表を**ヒストグラム**（Histogram）、その頂點を連ねた線を**度數多角形**といふ。

第一圖

第二圖

三　對數と對數圖表

統計計算に於てばかりでなく、日常遭遇する複雜な計算、例へば複利算などでは、對數は極めて偉力を發揮する。殊に對數の原理に據つて作られた對數圖表といふ特殊の統計圖表が廣く利用されるに至つて、對數に關する知識は一層必要になつて來た。よつて以下、多少詳細にこの原理と應用を說明しよう。

$y=a^x$なるとき、「xはaを底數とするyの對數なり」といひ、これを

$$x=\log_a y$$

と記す。但し $y=10^x$ なるときはこれを單に常用對數（Common logarithm）といひ、底數を記すのを省いて

$$x=\log y$$

と記す。底數は1以外の數ならば如何なる數でもよいが、一般には10か又はeを用ふる。eとは

$$1+\frac{1}{1}+\frac{1}{1.2}+\frac{1}{1.2.3}+\cdots\cdots$$ 卽ち

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$$

のことで、略々 2.71828183 である。この e を底數とする對數を自然對數といひ、純數學及び自然科學に於て廣大な應用範圍を持ち、經濟學や統計學の領域でも次第に應用されて來たが、こゝでは常用對數を説明しよう。

100は10^2であるから log 100=2 と記され、10は10^1であるから log 10=1 と記される。然らば100と10との間のある數、例へば50の對數は1から2までの間の或る數でなければならない。實際にこれを求めると $50=10^{1.6989700}$ 即ち log50=1.6989700である。次に1は10^0即ち log1=0 であるから——これは屢々間違はれ易いが、次の指數法則を理解すれば簡單な問題である——1から10までの間の或る數、例へば8の對數は0から1までの間の或る數であつて、實際にこれを求めると $8=10^{0.9030900}$ 即ちlog8=0.9030900である。

斯く或る與へられた數は、一般に10のx乘（即ち數學上の用語でいへば10の冪）と書き換へる事が出來る（但し與へられた數が負數及び零の場合は對數の理論は全く適用出來ない。蓋し10を何乘しても零又は負數とはならないからである。）故に $a=10^x$ であり、從つて loga=x である。

例へば

$$2=10^{0.3010300} \qquad \therefore \log 2=0.3010300$$

$15 = 10^{1.1760913}$　　$\therefore \log 15 = 1.1760913$

$133 = 10^{2.1238516}$　　$\therefore \log 133 = 2.1238516$

$1352 = 10^{3.1309767}$　　$\therefore \log 1352 = 3.1309767$

$25934 = 10^{4.4148695}$　　$\therefore \log 25934 = 4.4148695$

$0.2 = 10^{0.3010300-1}$　　$\therefore \log 0.2 = 0.3010300-1$

$0.05 = 10^{0.6989700-2}$　　$\therefore \log 0.05 = 0.6989700-2$

さて斯く一般に數字が對數で示されるとして、然らばそれでどんな結果を齎すか。これを知る爲には指數法則を明かにせねばならぬ。${}^{x}a$に於けるxをaの指數（Expo nent）といふ。この指數といふ文字は物價指數や生計費指數などの指數（Index number）とは全く別のものであることに留意されたい。

${}^{x}a$に於けるxなる指數は單にaをx度だけ掛け合はせる事を意味する。例へば

$a^3 = aaa$

である。この意味に於ける指數の性質を要約したものを指數法則といひ、次の如くである。

（一）　$a^m \times a^n = a^{m+n}$

例へば　$a^3 \times a^2 = aaa \times aa = a^5 (= a^{3+2})$

この事からして

$$\log (N \times M) = \log N + \log M$$

なる結果が生ずる。即ち　$N = 10^x$，　$M = 10^y$

とすれば　$N + M = 10^x \times 10^y = 10^{x+y}$となり、従つて

$$\log (N \times M) = x + y$$

然るに　$x = \log N, y = \log M$　であるから

$$\log (N \times M) = x + y = \log N + \log M$$

となるのである。故に例へば

$$\log (583 \times 2933) = \log 583 + \log 2933$$

（二）　$a^m \div a^n = a^{m-n}$

例へば　$a^3 \div a^2 = \frac{aaa}{aa} = a (= a^{3-2})$

この事からして

$$\log \left(\frac{N}{M}\right) = \log N - \log M$$

なる結果が生ずる。前例に從つて $N=10^x$ $M=10^y$

とすれば $\frac{N}{M}=\frac{10^x}{10^y}=10^{x-y}$

$$\therefore \log\left(\frac{N}{M}\right)=x-y=\log N-\log M$$

となるのである。故に例へば

$$\log(8533\div 952)=\log 8533-\log 952$$

である。更にこの法則によつて $10^0=1$ なる事は容易に證明出來る。即ち

$$10^0=10^{x-x}=\frac{10^x}{10^x}=1$$

だからである。この10の代りに如何なる數を代入しても矢張り同じである。即ち

$$a^0=a^{x-x}=\frac{a^x}{a^x}=1$$

となる。即ち零を指數とする數は常に一である。換言すれば $\log 1=0$ である。また

$$a^{-1}=a^{2-3}\ \frac{aa}{aaa}=\frac{1}{a}$$

即ち一般に

$$a^{-x}=\frac{1}{a^x}$$

即ち負數を指數とする數は、負數を正數に直した冪の逆數に外ならぬ。例へば

$$a^{-5}=\frac{1}{a^5}$$

この事が對數計算に如何なる關係を持つかといふに、例へば $0.2=\frac{2}{10}$ であるから

$$\log 0.2=\log 2-\log 10=\log 2-1$$

然るに $\log 2=0.3010300$ であるから、

$\log 0.2=0.3010300-1$ である。同様にして $\log 0.02=\log 2-\log 100=0.3010300-2$ である。更にまた、

$\log 20=\log(2\times 10)=\log 2+\log 10=0.3010300+1$ であり、$\log 200=\log 2+\log 100$

$=0.3010300+2$ である。

この 0.2, 0.02, 20, 200 の對數を見るに、何れも2の對數即ち 0.3010300 に $+2$ とか -1 とかが附加されてゐる事が判る。この 0.3010300 を假數といひ、$+2$ とか -1 とかを指標といふ。そして次の如く記す。

$$\log 200=2.3010300$$

$$\log 20=1.3010300$$

$\log\ 2=0.3010300$

$\log\ 0.2=\bar{1}.3010300$

$\log 0.02=\bar{2}.3010300$

斯くて對數に於ては、單に桁の違ふ同一數字は指標が變るだけで假數は變化しない事が判る。

例へば

$\log\ 58=1.7634280$

$\log\ 580=2.7634280$

$\log\ 5.8=0.7634280$

$\log 0.58=\bar{1}.7634280$

以上の事柄が判れば計算に於ける對數の意味は判然とする。即ちわれわれが與へられた數を對數に直して計算するのは必ず乘算及び除算に限られ、加算と減算に於ては用ひられないのである。10+5 又は 10−5 は、そのまゝ計算すればよく、もしこれを $\log 10+\log 5$、又は $\log 10-\log 5$ とすれば、前述の理由から 10×5 又は $\frac{10}{5}$ となつて了ふ。

對數計算には必ず對數表を用ひねばならぬ。對數表とは總べての正數の對數を表としたもので

ある。100の對數が2である事や1000のそれが3である事位ひならば常識からでも求められようが、8とか15とかの對數は普通の力では判るものではない。この計算の勞苦は數學者が引受けて驚くべき精密な對數表が出來上つたのである。對數表には四桁乃至五桁位ひの簡單なものから二十數桁に達する精密極まるものまで種々あるが、一般の用途には五桁乃至七桁のが便利である。桁數の多いものほど正確な答が得られるが、それだけ計算が面倒になる。いま七桁の表を用ひて簡單な計算を行つて見よう。

例一　$5329\times25.8\times586.69$　を計算せよ

求める數を x とすれば

$$\log x=\log 5329+\log 25.8+\log 586.69$$
$$=3.7266457+1.4116197+2.7684087$$
$$=7.9066741$$

これは求める數 x の對數であるから、$x=10^{7.9066741}$ である。指標が7であるから x が零を基準とした八桁の數字即ち千萬臺なる事は直ちに判る。その正しい答は對數表を逆に用い、如何なる數の假數が 0.9066741 なるかを求め、この數を八桁に改めればよい。これを求めれば

$x=80662954$

となる。計算器數を用ひ實際に計算すれば

$x=80662952.058$

となり、對數表による計算の結果は約2だけの誤差を示す。この誤差はより桁數の多い對數表を用ふれば次第に小さくなるが、八千萬に於ける2の誤差は先づ以て無視してよからう。

例二　3639^3 を求む

求める數をxとすれば

$\log x=3\log 3639=3\times 3.609821=10.6829463$

$x=48188800000$（實際に計算すれば 48188806119）

例三　$\sqrt[3]{59362}$ を求む

求める數をxとすれば

$=59362^{\frac{1}{3}}$

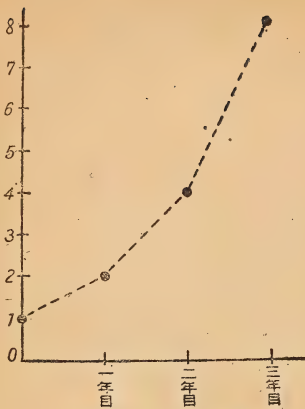
$\therefore\ \log x=\frac{1}{3}\log 59362=\frac{1}{3}\times 4.7778761\fallingdotseq 1.5926253$

$\therefore\ x=39.14$

例四　元金A、利率をrとすれば、n年後の元利合計は $A(1+r)^n$ であるから、$\log A+n\log(1+r)$ によつて求める事が出來る。蓋し $y=AB^x$ のyは $\log y=\log A+x\log B$ によつて求められるからである。例へば元金百圓、年利三分五厘とすれば、十五年後の元利合計 yは $y=100(1+0.035)^{15}$ であるから、兩邊の對數をとれば

$$\log y=\log 100+15\log(1+0.035)=2.2241045 \quad \therefore y=167圓53$$

普通のグラフに於ける縱軸及び横軸は何れも均分の目盛即ち算術目盛となつてゐるが、かゝる圖表は數列の絕對的變化を示すには適當でも、相對的變化を示す事は甚だ不便である。

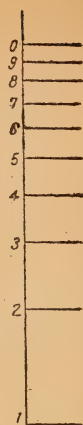
第　三　圖

假りに一年間に倍額になる割合で元金一圓が投資されたとすれば、一年後には元利合計は二圓となり、二年後には四圓となる。即ち増加率は同一であり乍ら、その絕對的増加額は年と共に増大する。これを圖示すれば上の如くである。（第三圖）。故に斯かる圖表を一覽したゞけでは増加率が一定かどうかは容易に判定し難い。これに反して

若し増加率が同一ならば、元本の如何に拘らず、常に直線で示されるやうな圖表が作られるとすれば、甚だ便利であらう。對數圖表とはこの目的に應ぜんとするものである。

上例の最初の金額一圓はその對數をとれば零となる（log1＝0）。そして一年後の元利合計即ち二圓のそれは 0.3010300 となり（log2＝0.3010300）、二年後のそれは 0.6020600 となる（log4＝0.6020600）。これを並記すれば次の如くである。

最初	￥1………0	
一年後	￥2………0.3010300	増加 0.3010300
二年後	￥4………0.6020600	増加 0.3010300
三年後	￥8………0.9030900	増加 0.3010300

即ちその對數を見れば、最初の一年間も次の一年間もその次の一年間も 0.3010300 だけ増加する事が判る。この理を應用し、絶對數の代りにその對數を目盛とすれば、増加率が同一なる限りは直線で示される筈である。……一圓の對數は零であるからこれを原點に置き、それから例へば一寸上に二圓の目盛を刻んだら、更にその上に同じく一寸の目盛を刻んでそれを四圓の目盛とするのである。この原理に從つて一から一〇迄の目盛を刻めば上の如くであるが（第四圖）その完

全なものは對數用紙又は對數方眼紙（Ratio-paper）の名で文具店で賣つてゐる。なほ縱軸だけを對數目盛としたものを半對數圖表（Semi-logaritimic chart）といひ、時の經過に伴ふ變化の割合を示すに用ひられる。普通にいふ對數圖

第四圖

表とは殆ど常にこれを指す。然るに時には縱横ともに對數目盛にしたものがある。これは横軸に時以外のものを取つたときに屢々使用され、例へばパレートの所得分配線の如く縱軸に所得税納入者の數を、横軸に所得金額を示す場合には、兩者が對數目盛となつてゐれば甚だ便利なのである。併しそれを說明するのは相當の紙數を要するので、此處では割愛せねばならぬ。

半對數圖表のみについていへば、その主たる目的は時の經過に伴ふ增減の割合を簡單且つ的確に示すに在る。併し目盛と照し合はせれば絕對量を讀む事も可能であるから、算術圖表の役目をも兼ねてゐるといへる。半對數圖表に示される線の角度は增加率又は減少率によつて決定されるから、圖表を一瞥して直ちに增減の比率を推知し得るのである。即ち第五圖の示すが如く、

（一）線が底線に平行なる時は量に增減なきを示し（次圖のA線）

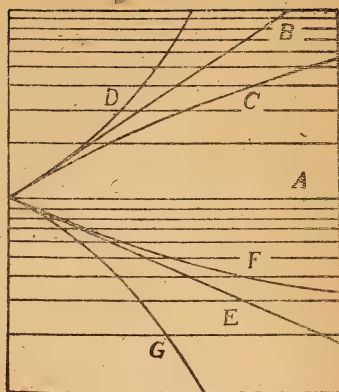
（二）上昇的直線は量が一定比率にて增加するを示し（B線）

（三）右に行くに從つて角度の減ずる上昇線は、量は增加しつゝあるも其の率は漸減しつゝあるを示し（C線）

第五圖

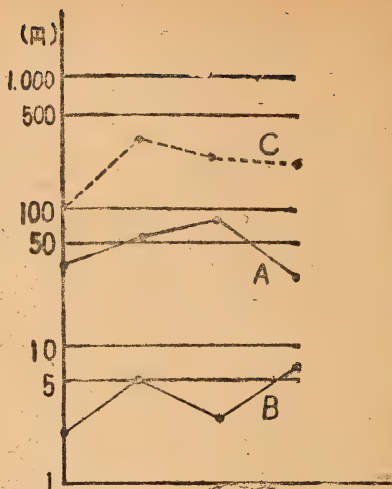
（四）右に行くに從つて角度の增す上昇線は、量も增加率も共に漸增しつゝあるを示し（D線）

（五）下降的直線は量が一定比率にて減少しつゝあるを示し（E線）

（六）右に行くに從つて角度の減ずる下降線は、量は減少しつゝあるも、減少率は漸減しつゝあるを示し（F線）

（七）右に行くに從つて角度の增す下降線は、量が減少すると共に減少率は漸增しつゝあるを示す（G線）

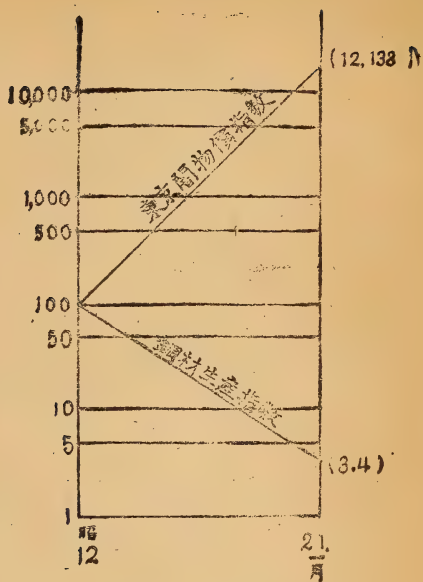
（八）同一圖表に示された二線が、平行なるときは增加又は減少の率が同一なるを示し、平行ならざるときは、急峻な線はより大なる增加率又は減少率を示す。

對數圖表の一つの長所は、圖表の上で簡單に掛算または割算の計算が行へることである。生産

量Aと單位價格Bがこの圖表の上に描かれてゐれば、生産價額C即ち A×B は圖表の上ではAの高さにBの高さを加へることによつて求められる。逆に、生産價額と生産量とが與へられてゐれば、單位價格は前者を後者で割ることによつて、即ち前者の高さから後者の高さを引くことによつて求められる。コンパスがあれば誰にでも計算は容易である。

第六圖

(第六圖)

對數圖表のもう一つの利點は、一つの表に極めて大きな數字と小さな數字とがある場合、これを比較的小さなグラフで示せるといふことである。次のグラフを普通目盛のグラフにすれば何尺にも延びて了ふであらう。(第七圖)

10,000
5,000
1,000
500
100
50
10
5
1
(12,138)
(3.4)
東京闇物價指數
鋼材生産指數
昭12
21.1月

第　七　圖

第四章　統計的比率

一　絶對數と比率

合理的判斷の基礎が比較であり、そのために數字が必要なことは、既に冒頭に述べたところである。對象が數字化されゝば、その數字を比較することによつて容易に且つ正確に大小、强弱、優劣、增減等々即ち一言にしていへば程度の相違を判斷することができる。このことは對象が個體でなくて集團のときでも同じで、その場合にはわれわれは統計を比較すればよいのである。統計が集團的事實を正確に記述する手段たる所以はこゝに在る。では數字の又は統計の比較はどうすれば行へるか。一般に比較は簡單な問題と考へられ易いが、實は統計學の本質に觸れた最も困難な問題で、統計學を一應卒業した後でなければ詳しくは扱へない性質のものなのである。こゝでは基本的問題を略述するに止めざるを得ない。

數字の比較とは二つの數字の何れがどれだけ大きいかを見ることであるから、いま二つの數字をx及びyとすれば、

(1) x—y,　または　y—x

(2) x/y,　または　y/x

の何れかの形で行はれる。前者は差を、後者は比（比率または比例數ともいふ）を見ることである。差の場合にはどちらからどちらを引いても絶對値は變らないが、比の場合にはどちらを分母とするかで値は變つてくる。5に對する10の比なら2、10に對する5の比なら0.5である。一體差または一般に絶對數と比と何れが重要かと聞かれると答へる術がない。問題によつてその都度きめる外はないからである。二三の場合を考へて見よう。食糧政策の問題としては食糧がどれだけ現實に不足してゐるかゞ先決問題であるから必要量と生産量或ひは供出量との差が重要である。それによつて必要な輸入量とか可能な配給量とかゞ決定されるから。併しその場合の一人當り配給量は總量を人口で割ることによつて、即ち比率の式で求めねばならない。學問的には、必要量と生産量との比による所謂自給率がより問題となることがある。併しこの問題で最も適當な例は恐らく人口増加に關する事柄であらう。周知の通り、戰前戰時を通じて我國の人口政策は所謂「産めよ殖えよ」一色に塗りつぶされた。その論據は大正元年以來の出生率遞減傾向に求められたのである。事實同年の出生率（人口千人に對する出生數）は三六・一九で、從來の最高記録を示

した。その後緩慢ながら次第に低下し、昭和十年前後には略々三〇、日華事變の開始翌年には二六・七〇に降つた。人口政策確立要綱といふ極端な増殖案が決定されたのは、この低下に驚愕したからである。ところが出生を比率としないでその絶對數を見れば、大正九年の出生總數は二、〇二五、五六四、昭和十年前後のは大體二一〇萬で、逆に少からず増加してゐる。この絶對數と比率との逆關係は勿論基礎人口の増加による現象であるが、何れにしろ比率だけを問題にして絶對數を見ない態度は公平とはいへないのである。事實その後の我國の人口動態は死亡の著しい改善によつて（大正九年には一四二萬强、昭和十年前後には約一二〇萬、從つて死亡率は大正九年には二五强、昭和十年前後には一七乃至一八）、現實の一年間の人口増加は大正九年の六〇萬强が昭和十年には一〇〇萬を突破した。既に過剩人口に惱む國に於ては、問題は明かに増加の絶對數であつて、比率は二次的意義しかもたない筈である。出生率は勿論、増加率さへ假令いかに減つたところで、現實の人口が殖えてゐれば、密度は益々大となり、他の事情が等しい限り、過剩度は益々甚だしくなるのである。單に増殖論を裏づけるために、比率を過當に重視し絶對數を不當に閑却した學者のいかに多かつたことか。

同じことが、カール・マルクスの産業豫備軍說についても言へる。彼に從へば資本は、機械や

原料や設備等に充用される不變資本と、勞働雇傭に充用される可變資本とに分たれる。資本主義的生産方法の進歩と共に前者は相對的に増大し後者は相對的に減少する。彼はこれを資本の有機的組成の高度化とよんだ。そこで勞働者は機械に追出されて益々失業し、山積されて産業豫備軍といふ尨大な失業者群團（マルクスの所謂資本に對する相對的過剩人口）ができ上る。彼等は一般賃銀の上昇を阻む作用をなし、勞働階級の貧窮を一層甚だしからしめるが、かゝる群團が更に膨張して人口の大部分を占めるに至れば、社會の購買力は低下し、他方の益々高まる生産力との不均衡は最後に全面的恐慌となつて爆發し、資本主義そのものが終熄するといふのである。いかにも尤もらしいが、併し彼の前提からは必ずしも産業豫備軍は發生しないのである。假りに勞働人口も總資本も共に不變とすれば、有機的組成の高度化は必ず勞働者を驅逐する。併しマルクスはこの兩者の増加を前提としてゐる。然らば可變本資は、不變資本ほど速かに増加しなくても、絕對的には増加しうるのであつて、もし總資本の増加が甚だ大で、且つ有機的組成の高度化は緩慢にしか行はれないとすれば、可變資本のこの絕對的増加は増加的勞働人口を完全に吸收して、場合によつてはなほ餘りがあらう。總資本と勞働人口がどれだけ増加するか、及び資本の有機的組成の高度化がどれだけ進行するか、問題はこゝに懸るのであつて、單に比率だけの問題として

は解決され得ない。私は産業豫備軍の發生に關するマルクスの眞意は正しいと信じてゐる。こゝでは唯だ、彼の前提は不充分だといふのである。（この問題については拙著「人口理論の展開」を參照されたい。）

比率だけで問題が解けないと同樣、絕對値だけでも不充分である。例へば人口の眞の增加力を測定するためには、比率が決定的役割を演じなければならない。出生數の激減した歐洲諸國でも、二三の例を除けば、出生はなほ死亡を超過し、從つて人口は年々幾分づゝ殖えてゐる。人口動態の最惡狀態に陷つた一九三三年のドイツでも、出生數は九六萬弱（比率で一四・七）、死亡數は七二萬弱（比率で一一・二）、從つて約二三萬（比率で三・五）だけ現實に增加してゐる。勿論自然增加が九〇萬を突破した今世紀初期と較べれば、增加數のこの激減は驚くべきものがあるが、兎に角現實に增加してゐることは事實であるから、素人の眼には、人口增加力は現在人口を保持して餘りあるものゝ如く映つても不思議はない。併しこれは一種の幻覺に過ぎない。ドイツでは今世紀初頭以來の出生率の直線的低下によつて幼少年人口の割合が甚だ少い。故に壯年者は、特別多く子供を產むわけではないが、總人口に比しては割合に多い、即ち出生率は比較的に高い。このまゝでゆけば、現在子供を產んでゐる人口はどんどん老人になり、子供を產まないと

共に死亡も殖える。而も後に續く若年人口は少いから、彼等から澤山の子供が生れる筈はない。出生は急に減り、死亡は殖え、人口增加は逆に人口減少に變るであらう。即ち現在の增加は單に人口の年齡構成の上から起つてゐる過渡的現象に過ぎないのであつて、眞の增加力は現在人口を維持するに足りない極めて微弱なものなのである。謂はゞ簡生活によつて外見をつくろつてゐるようなもので、賣るものがなくなれば、收入不足は覿面でなければならない。

この場合には要するに表面の絕對數では眞相は判らない。人口統計學者はこれを正確に測定するため、純再生產率という特殊の比率を計算する。女子人口について年齡別出產率を求め、これを年齡別死亡率で修正して綜合したもので、これによつてわれわれは、現在の一人の母が次代に幾人の母で置換へられるかを知ることができる。純再生產率が一ならば、現在の一人の母は次代に一人の母で置換へられ、一・五ならば一・五人、〇・八ならば〇・八人で置換へられることを意味する。一九三三年のドイツのそれは一を遙かに下廻つてゐて、表面上の增加とは逆に、眞の增加力は甚だ心細い狀態に在つたことが窺へるのである。

二　比率の種類

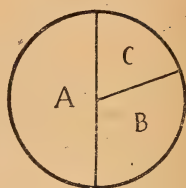
比率は一方を分母、他方を分子とする單純な形であるが、內容的には次の三つの種類がある。

(一)構成比率(分析的比率)　(二)對級比率(指數)、(三)混成比率(關係比例數)。

(一) 構成比率

第一の構成比率とは一つの全體が幾つかの部分から成るとき、各部分が全體に對して占める割合のことである。A＝x＋y＋z ならば、構成部分即ち x y 及び z のそれぞれの割合は $\frac{x}{A}$・$\frac{y}{A}$・$\frac{z}{A}$ である。三者の合計は一であるが($\frac{x}{A}+\frac{y}{A}+\frac{z}{A}=\frac{x+y+z}{A}=\frac{A}{A}=1$)、一般には一〇〇倍して%で示す。即ち合計は一〇〇となる。構成比率は事物の構造狀態を示すもので、謂はゞ分析の結果を示したものであるから、人によつてはこれを分析的比率と呼んでゐる。その用途は殆ど無限で、われわれは到るところでこの應用例に遭遇する。溫泉浴場に飾られた分析表、賣藥の効能書にある調劑內容、驛のポスターに記された事故の割合等々、數へればきりがない。

生活水準の判定標準としてエンゲル法則といふものがある。貧乏人ほど生計費の中で食費の占める割合は高いといふことで、これは食費の構成比率の如何を言つてゐるのである。食物は絶對必要品で、經濟學的にいへば需要彈力性の最も乏しい品であるから、他の支出ほど節約することはできない。極端な貧乏になれば、有金全部を食物に當てて了ふであらう。戰前我國の勞働階級

は食費に所得の三〇%前後しか使はなかつたが、今日では六〇%から七〇%に達してゐる。所得それ自體は五六十倍に上り乍ら、その大部分は食費に喰はれて了ふわけで、別の言葉でいへば、名目賃銀は上り乍ら、實質賃銀は逆に低下したことを物語つてゐる。このことは一國全體の收支狀態に就てもいへることで、國民所得が少ければ少いほど、文化的支出と比較して食糧に對する支出は多からざるを得ない。即ちエンゲル法則は個人のみならず一國の貧富を判定する恐らく最良の規準であらう。構成比率はこれをグラフで示すのが最も見易い。いま或る商品の生産費を分解したら、生産要素A、B、Cのそれぞれの割合は50、30、20だつたとする。その圖示方法は構成棒圖表かパイ圖表による。前者は一本の棒を右の割合に分割すればよい（第八圖A）。後者は一つの圓をその構成要素の割合に從つて中心から扇形に分割する方法で、パイ（pie）といふ洋菓子を切るときの形に似てゐるため、パイ圖表の名がある。中心角は360°であるから、1%を3.6°の割合で計算すればよい。即ち $50\times3.6^\circ=180^\circ$，$30\times3.6^\circ=108^\circ$，$20\times3.6^\circ=72^\circ$ となる（第八圖B）。

二つの半徑によつて作られた面積は、中心角にも弧にも比例するから、面積、中心角または弧の何れを見てもよいわけである。

(A)　(B)

第八圖

(二)　指　數

同種の數字を對立的に比較したもので、一方を基準（一〇〇）とし、他方を基準に對する比率（%）としたものである。物價指數、生計費指數、生産指數、賃銀指數、景氣指數等々極めて廣い範圍に亘つて利用され、統計値として最も重要なものの一つとなつてゐる。併し右の各種指數は所謂綜合指數で、その理論及び方法は甚はだ複雜であるから、説明は後章に讓るとして、こゝでは最も簡單な個別指數に限定しよう。

同種の數字 a_1 a_2 a_3……a^n が與へられたとき、例へば最初の a_1 を基準とすれば、a_2 a_3 等々は $\frac{a_2}{a_1}$, $\frac{a_3}{a_1}$ ……と書換へられる。實際には a_1 を一〇〇とし、從つて以下はそれぞれ一〇〇倍して $\frac{a_2}{a_1}\times100$, $\frac{a_3}{a_1}\times100$……とする。構成比率は各要素が集つて一つの全體を構成してゐる場合で、指數は各要素が互ひに獨立し、唯だ種類が同じなため一つの數列として竝べられてゐる場

合である。但し注意しなければならぬことは、各要素が集つて一つの全體を構成してゐるときでも、見方によつては指數で示すことも出來るといふことである。前例の領土の問題で、日本の領土を基準とし、朝鮮・台灣等々のそれを基準に對する%とすれば、一つの指數になる。

領土	日本	朝鮮	台灣	………
面積	383	221	36	………
指數	100	57.7	9.4	………

何が同種で何が異種かは、既に述べた通り、場合によつてどうともなるから、構成比率の大部分は指數でも示せるわけである。尤もどちらがよいかは、問題の性質で自ら決まらう。反之、指數を構成比率に改めることは、多くの場合不可能である。といふのは、指數が最も屢々作られるのは時系列即ち時の經過に伴ふ變化を示した統計であるが、かゝる統計には全體といふ綜合物がないからである。全體がなければ總和がなく、總和がなければ總和に對する割合もない。

特定商品の價格變化は個別指數で示される。これを價格指數といつて、綜合現象たる物價の變化を示す物價指數と區別する。基準時點の價格をp_0、比較時點のそれをp_1とすれば比較時點の價格指數は$\frac{p_1}{p_0}$である。個別指數での問題は基準をいかに決めるかといふことだけで、綜合指數の

ように加重や平均の難問に惱まされることはない。基準のとり方は双方に共通した問題である。基準は素より出來るだけ正常的なものでなければならぬ。これは凡ゆる比較に共通の眞理である。人の賢愚優劣は普通人と比較していふべきで、偉人や白痴と比較しても意味をなさない。價格指數の作製に際して、何等か異常な原因で價格が異常に高かつた時を基準とすれば、指數は總べて小さな値となり、反對に、價格が異常に低かつた時を基準とすれば、大きな値となる。別段デフレもインフレも起つてゐないのに、指數の表面ではかような疑ひをもたせることにもなる。併しこの問題も綜合指數の章まで保留しよう。

注意しておきたいことは指數といふ文字のことである。第一に、こゝにいふそれは數學上の用語たる指數とは別物だといふことである。後者は a^x に於ける x（即ち冪を示す指數）のことで、英語では Exponent といひ、こゝでいふ指數は Index-number といふ。日本語では同じ文字で異るものを指してゐるわけで、このことは對數に關聯して既に述べたことだが、不便と言はざるを得ない。第二に、生物學や醫學で用ひられる頭蓋指數や鼻高幅指數などに於ける指數は、次に述べる對立比率に屬し、こゝに謂ふ指數とは別物である。頭蓋指數とは頭の長さに對する頭の幅の割合で、鼻高幅指數とは鼻の高さに對するその幅の割合である。高さや長さと幅とは

異種の數字であるから、こゝでいふ小指數とは全然ちがつたものである。この種の計測指數は人體に關するものだけでも甚だ多く、羽田宣男氏著「生體計測、人類學の基礎」には九十種も擧げてある。

（三）　混成比率（關係比例數）

分子と分母とが種類を異にする場合であつて、關係比例數ともいふ。異種といつても全く無關聯の數字を對立させたのでは計算の意味がなくなるから、兩者の間に何等かの聯りがなければならぬことは言ふ迄もない。この聯りには自然的なものと社會的なものとがある。自然的關聯とは分子と分母とが因果的關係即ち分子が分母から發生する關係に立つものを指し、社會的關聯とは兩者が單に相關的關係即ち分子の意味が分母との比較によつて始めて社會的に確定されるものを指す。前者を發生比率、後者を對立比率ともいふ。人口と、それから發生した出生數、死亡數または婚姻數との比較によつて得られる出生率、死亡率または婚姻率とか、資本と、それから生じた利潤との比較によつて得られる利潤率とかは發生比率である。面積と人口との比較による人口密度や、交換で授受される二商品の量の割合即ち交換比率は對立比率である。

混成比率に於ける最も困難な問題は、分母と分子との關聯を如何に規定するかといふことであ

る。例へば出生は總人口から發生したものではあるが、併し總人口の中には幼少年又は高齢のため生殖には無關係の人口も含まれてゐる。更に、直接子供を生むのは婦人であり、特に有配偶婦人である。故に出生と對比さるべき人口は、總人口、生産年齢人口、姙孕可能年齢婦人々口、有配偶の姙孕可能年齢婦人々口等がありうるのであつて、從つて單に出生率といふ場合その何れを指すかは判然としないわけである。一般に分母が分子に直接關係のない要素をも含む場合を粗率といひ、これらを排除した場合を精率といふが、それらにも種々の段階がありうることは、いま述べた例から容易に判るであらう。そして嚴正を期するにつれてより高次の精率が必要となるのである。このことは純再生産率の問題に觸れて既に一言した。

第五章　數値の要約（平均の理論）

一　平均値の概念

森に入つて森を見ずといふ諺通り、細部に拘泥すると大局の見透しはつかないものである。ところがかやうな見透しが必要だといふことが、統計を生んだ抑もの理由であるから、統計的結論は常に、簡單な數字で能く全體を表現できるものでなければならない。さて度數系列とは全單位をば數字的分類基準に從つて幾つかの級に分けたものである。分類とは元々複雜なものを見易い形に改める一つの手段であるから、度數系列はそれ自體既に要約的性質を帶びたものである。一人一人數へれば一億ある人口も、各歳別に分類すれば百ぐらゐの級になつて了ふ。ところがこの場合、いくつに分けるか卽ち級間隔をどの位の大さにするかについては、別に規定はない。年齢分類には各歳別も、五歳別も、十歳別その他もある。餘り細かく分けたのでは、結果たる度數系列は長つたらしい見にくいものになつて、要約といふ要請に合はなくなる。では逆に級間隔を大きくして級の數を少くすればよいかといふに、必ずしもそうではない。年齢別で全人口を零歳か

ら四十九歳まで、及び五十歳から百歳までとなつた二つの級に分てば、分類の効果は殆どなくなり、もし更にこれを壓縮して零歳から百歳までを一つの級とすれば、全く分類が消滅するわけで、同時に要約の効果もなくなつて了ふ。われわれの要求するものは斯かるものではなく、より簡單な卽ち幅をもたない或る一つの數字で能く全體を示しうる方法である。

いま甲乙二組の學生があり、甲組は一八歳、二〇歳、二六歳の三人から成り、乙組は一六歳、一八歳、二二歳、二五歳の四人から成るとする。どちらの組がより若いか。これは分類しても判らず加へ合せても判らない。加へ合せれば甲組は六四歳、乙組は八一歳だが、一方は三人、他方は四人だから、甲組の方が若いとは言へない。そこでこの合計を人數で修正し、一人當りの年齢を求めて見ると、甲組は二一・三分一、乙組は二一・四分一となり、乙組の方が少し若いことが判るのである。一人當りの値は卽ち代表値であり從つて要約値である。われわれは與へられた系列の代表値を求めることによつて、その系列を要約しうるのである。

いま右の代表値を見るに、それは一組の中に含まれた若い者と年長の者との中間の値である。中間とは二つの極端の間といふことで、これを中央といつてもよい。併し乍ら中央とは何かといふことになると實は客易ならざる問題である。百里の中央は五十里かといふに必ずしもそうでは

ない。九十里を以て半ばとする人もあり、また、よく始められたものは既に半ば成就されたに等しいと嘯いて、最初の一里が既に半ばだと主張する人もあらう。問題の置き方または見方の如何で中央といふ文字はどうとも解釋されるのである。して見ればそれに應じていくつもの平均値がなければならぬ。

二　平均値の種類と性質

斯くて我々は幾通りかの平均値を考へうるのであつて、考へ方の僅かの相違をも考慮に入れゝば、その種類は無限と言つても過言ではない。併し一般に用ひられるのは四乃至五種類に過ぎない。平均値は勿論數字的にも色々の問題があるが、數字的事實を取扱ふ統計學に於ては正に中心的課題であり、統計學者のうちには統計學を以て平均の學と定義した人すらある。以下、主たる平均値とその性質の一般を述べよう。

(一)　算　術（相加）平均

一數列（變數）を構成する各項（變量）の値を加へ合せ、これを項數で除した商をいふ。即ち變數Xが x_1　x_2　x_3……x_n なるn箇の變量より成るとき、その算術平均mは

$$m=\frac{1}{n}(x_1+x_2+x_3+\cdots\cdots+x_n)=\frac{1}{n}\Sigma(x)$$

例へば5円8円11円のそれは、$\frac{1}{3}(5+8+11)=8$円 である。斯く箇々の變量が與へられてゐるときは至極簡單であるが、全項がいくつかの級に分類されて度數系列の形をとるときは、度數を算入せねばならぬから、次の形となる。

$$m=\frac{1}{\Sigma(f)}\Sigma(f\cdot x)$$

この式のfは度數、xは級の中點を示す。次表は二六一人の工員の週給が四圓づゝの級に分たれてゐるときの計算例である。

週給別職工

階　級	中點	度數(f)	f×中點
¥4—7.99	¥6	5	30
8—11.99	10	15	150
12—15.99	14	46	644
16—19.99	18	68	1224
20—23.99	22	58	1276
24—27.99	26	32	832
28—31.99	30	22	660
32—35.99	34	10	340
36—39.99	38	2	76
40—43.99	42	2	84
44—47.99	46	0	0
48—51.99	50	1	50
		261	5366

$$m=\frac{5366}{261}=20.56$$

但しこの例のように級間隔が等しいときは、最初の中點（六圓）を基準とし、級間隔（四圓）を一單位として書換へれば、計算は甚だ容易になる。これから平均値を求める場合には、950 に四圓を乗ずることゝ、それを基準の六圓に加へることを忘れてはならぬ。

x	x'	f	$x' \times f$
6	0	5	0
10	1	15	15
14	2	46	92
18	3	68	204
22	4	58	232
26	5	32	160
30	6	22	132
34	7	10	70
38	8	2	16
42	9	2	18
46	10	0	0
50	11	1	11
		261	950

$$m = 6 + \frac{950}{261} \times 4$$

$$\fallingdotseq 20.56$$

また算術平均値と各項との偏差の合計は零に等しいといふ下記の定理を利用して、次のような計算もできる。これは任意の値（こゝでは二二圓）を假定平均値とし、これを基準として且つ級間隔四圓を單位として偏差を求めたもので、もし假定した二二圓が眞の平均値であつたとすれば、正の偏差と負の偏差は相等しく、從つて合計は零とならねばならぬ。ならぬ場合はなるように修正すればよい。即ちこの例では −94 だけ餘つたから、これに級間四圓を乘じて元の單位に戻し、

これを假定平均二二圓に加へればよいのである。この方法だと運算は最も簡單で、多少複雜な資

x	x'	f	$x'\times f$	
6	−4	5	−20	
10	−3	15	45	−225
14	−2	46	−92	
18	−1	68	−68	
22	0	58	0	
26	1	32	32	
30	2	22	44	
34	3	10	30	131
38	4	2	8	
42	5	2	10	
46	6	0	0	
50	7	1	1	
		261	−94	

$$m=22+\left(\frac{-94}{261}\times 4\right)$$

$$=22-1.44=20.56$$

料については常にこの方法を採るべきである。

これら計算に於て、必ず級の中點をとる理由を一考しよう。度數系列に於ては、全項の値の總計は不明である。算術平均を適用するにはこの總計値が絕對に必要であるから、右表の如く各級の中點をその級の代表値と假定するのである。即ち初項は、五人の工員の週給は四圓乃至八圓といふ意味であるから、その合計は 5×4円→5×8円 即ち二十圓以上四十圓以下といふこととしか判らない。そこで中央の六圓を以てこの五人の各々の週給と假定し、從つて合計を 5×6円 即ち三

〇圓と假定したのである。その妥當性は畢竟確率理論から求められるのであつて、中央値は所謂最確値と認められたからである。他の級についても亦同じ。斯かる假定によつて、全員の週給總計は五三六圓となり、よつてこれを度數の總和即ち全工員數で除すことによつて一人當りの賃銀即ち算術平均値を求めるのである。

算術平均値の性質は次の如く要約される。

（1） 算術平均は變量の總和を變量の數で割つた數であるから、相加平均と項數との積は全項の總和に等しい。即ち $n\times \Sigma M=(x)$。 斯くてその構造は後に述べる期望値（確率の章參照）と同じである。

（2） 算術平均と各項との差（偏差）の總和は零である。即ち $(x_1-M)+(x_2-M)+\cdots\cdots+(x_n-M)=0$ である。その證明、

$$(x_1-M)+(x_2-M)+\cdots\cdots+(x_n-M)=\Sigma(x)-nM$$

然るに（1）によつて $nM=\Sigma(x)$ であるから、$\Sigma(x)-nM=0$ これを利用した簡便計算法は既に例示した。

（3） 算術平均と各項との偏差の自乘の總和は、他の如何なる値から測つたそれよりも小さい。即

ち他の値をAとすれば

$\Sigma(x-M)^2 < \Sigma(x-A)^2$ である。その證明。

$\Sigma(x-A)^2 = \Sigma(x-M+M-A)^2 = \Sigma(x-M)^2 + 2\Sigma(x-M)(M-A) + (x-A)^2$

然るに右邊第二項$\Sigma(x-M)_2$は（1）によつて零であるから、第二項そのものも零である。そして第三項は $M=A$ のときは零、その他の場合は正であるから、右邊の總和は $M=A$ のとき、即ち任意の値が相加平均に一致したとき、最小である。後に述べる最小自乘法とはこれに立脚する命題である。

（4） 二數a、bの算術平均はその等差中項に等しい。

即ち $M=\frac{1}{2}(a+b)$　　$\therefore (a-M)=(M-b)$

（5） 算術平均値は物理學上の重心或ひは均衡點を意味する。

挺子の理論によれば、O點に働く力（能率（モウメント））は、加へられた力と距離との相乘積である（第九圖）。いま $f_1 \times a = f_2 \times b$ ならばO點に於ける力は相均衡し、挺子は水平に保たれる。f_1は四キロ、f_2は六キロ、挺子の長さは50糎とすれば、O點は $a=30$ $b=20$ とした點である。蓋し

$4\times30=6\times20$

なるを以てゞある。然るにいま一〇圓の品が四箇、五圓の品が六箇あれば、平均値は

$$\frac{10\times4+5\times6}{4+6}=7$$

第九圖

即ち七圓である。この七圓は一〇圓と五圓との幅を三對二の割に區分する點で、挺子の0點に等しい。換言すれば七圓を基準とすれば、$a=10-7=3$ $b=5-7=-2$ となり、$(4\times3)+(6\times-2)=0$ を得る。これは偏差の總和が零なることを意味する。

註 支點をa、力の作用線までの距離をx、力の大きさをyとすれば、xyを以て支點aに關する力の**モーメント**(moment)といふ。aはいかなる點であつても差支へないが、上記の如く算術平均値を選べば、均衡狀態が得られる。いくつの力がいくつの點に働いても正負のモーメントはこの點で相殺される。即ち $x_1y_1+x_2y_2+\cdots\cdots=\Sigma(xy)=0$ である。距離を二乘すれば二次のモーメント、三乘すれば三次のモーメントと呼ばれる。一次、二次、n次のモーメントをそれぞれμ_1 μ_2 μ_nと記す。統計學に於ては極めて重要な役割を演ずる。

(6) 算術平均値は變量の總計値と項數だけが與へられゝば計算できる。反對に總計値が不明な場合は計算できない。統計には屢々一端または**兩端の開いてゐるもの**がある。例へば六〇歳以上が

何人として與へらるゝが如き之であつて、この場合には六〇歳以上の人々の年齢合計不明であり、從つて全人口の年齢合計が不明となるのである。併し乍ら斯かる階級に含まれる度數は少いに決つてゐるから、實際には無視できる場合が多い。但し所得統計に於て最低所得者を「年收千圓以下何人」と記した場合には、その人數は極めて多いから、右の言は適用できないであらう。

(二) 幾何平均

一數列を構成する各項の値を掛け合せ、これを項數で開いた開冪をいふ。即ち

$$G = \sqrt[n]{x_1 x_2 x_3 \cdots x_n} = \sqrt[n]{\pi(x)}$$

度數系列については

$$G = \sqrt[\Sigma(f)]{X_1^{f_1} \cdot X_2^{f_2} \cdot X_3^{f_3} \cdots \cdots X_n^{f_n}} = \sqrt[\Sigma(f)]{\Pi(xf)}$$

これが計算には對數に書改める必要がある　即ちそれぞれ次の通りになる。

$$\log G = \frac{1}{n}(\log x_1 + \log x_2 + \cdots\cdots + \log x_n) = \frac{1}{n}\Sigma \log x$$

$$\log G = \frac{1}{\Sigma(f)}(f_1 \log x_1 + f_2 \log x_2 + \cdots + f_n \log x_n) = \frac{1}{\Sigma(f)}\Sigma f \log x$$

幾何平均の性質

(1) 二數の幾何平均値はその二數の等比中項に等しい。即ちそれは二數を比例的に均衡せしめる値である。一〇圓の價値あるものを甲は二〇圓、乙は五圓と評價したとすれば、兩者の犯した誤謬は等しいと見るべきであらう。蓋しいかに低く評價しても零が止まりであるに對し、高い評價には際限がなく、兩者の犯しうる誤謬の範圍は決して等しくないからである。零に對應する値、即ち零と相殺される値は、算術平均ならば二〇圓であるが、この問題に於ては無限大でなければならず、從つてこの理を押進めれば、幾何平均をとらねばならぬ。要するに偏差が幾何級數的と考へられるものについては、幾何平均によるのが正當だといふことになるのである。

(2) 右の理由から、同一變化率を以て變動すると假定された現象の平均値は幾何平均によつて求めらるべきことが判る。Aなる人口がn年度にBとなつたとし、均一増加率を假定すれば、平均人口は $(A+B)/2$ ではなくて $\sqrt{AB}$ である。Aを二萬、Bを八萬とすれば、平均人口は五萬ではなくて四萬である。その増加率は $B=A(1+R)^n$ から

$$1+R=\sqrt[n]{\frac{B}{A}}$$

として計算される。一年の年頭と年末の人口は異るから、その年の人口とは普通は兩者の平均を指すが、右の理由からそれは幾何平均による。

經濟學及び統計學に於ては比率を取扱ふ場合が極めて多いから、幾何平均の應用範圍も自ら極めて多いわけである。幾何平均の重要さはジェヴォンスによつて始めて强調され、物價指數に關してはその適當なる所以は既に明かにされた。併し乍ら今猶ほこの平均法が不當に閑却されつゝあることは多くの學者の指摘するところである。

註1　同一資料について算術平均と幾何平均を求めれば、後者は必ず前者よりも小さい。

稀には**調和平均**（Harmonic mean）なるものが使用される。調和平均とは「各項の逆數の算術平均の逆數」である。即ち

$$H=\frac{1}{\frac{\frac{1}{x_1}+\frac{1}{x_2}+\cdots\cdots+\frac{1}{x_n}}{n}}=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots\cdots+\frac{1}{x_n}}$$

（三）中位數（メディアン）（median）

一數列を構成する各項を大さの順序に列べ換へたとき、中央に來た項の値をいふ。項數が偶數の場合には、中央項はないが、その場合には中央二項の算術平均値を以て之に代へる。但し項數が極めて多いときは、中央二項の何れかの値で代用しても差支へない。度數系列については、度數

を順次に加へ合せて累積度數を作り、以て中央項の屬する階級を求め、次で補間法によつて中位數を決定する。

階級	度數	累積度數
¥4—7.99	5	5
8—11.99	15	20
12—15.99	46	66
16—19.99	68	134
20—23.99	58	192
24—27.99	32	224
28—31.99	22	246
32—35.99	10	256
36—39.99	2	258
40—43.99	2	260
44—47.99	0	260
48—51.99	1	261
	261	

$$Me=L+\frac{i}{f}C$$

$$=16+\frac{65}{68}\times 4$$

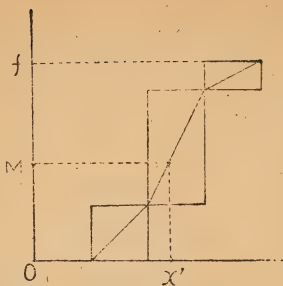
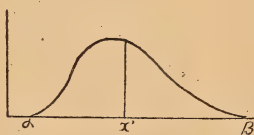
$=1$[illegible]$.82$

算術平均を求めたときの例について、中位數を計算すれば右の如くである。全數が二六一人であるから、中位數は賃銀順に列べた一三一人目の賃銀である。累積度數によつて、それが第四級に含まれてゐることは明かであるから、中位數は16円+xである。この一六圓をLで示す。一三一人目は六八人より成るこの第四級の第六五人目に當るから、級の大さ四圓を65:68に分つたものがxに該當する。右式のfは六八人、iは六五人目を意味する。

中位數はグラフの上からも求められる。簡單のため級の數を三つとすれば、次圖の如くそれを

順次に積み重ね、その全體の高さofを二等分する點Mからx軸に引いた平行線と、棒圖の對角線との交點を求めれば、この交點からx軸に下した垂線の足x'が求める中位數である。（第十圖）

理由は格別説明する迄もあるまい。なほ度數分布が度數曲線 $y=f(x)$ となれば、解析的には中位數x'はこの曲線の包む面積を二等分する點として、次式で示される。（第十一圖）

$$\int_{\alpha}^{x'} f(x)dx=\int_{x'}^{\beta} f(x)dx$$

第十圖

第十一圖

平均値とは中庸の値とも解し得よう。いま大さの順に列べられた一系列から先づ兩極端の項を取去り、次に次の兩極端項を取去るといふ風にしてゆけば、最後には眞中の一項（偶數項ならば

二項）が殘る。これは極端といふ要素は全く持たぬところの中庸項に外ならず、從つてその値たる中位數は正しく一箇の平均値である。中位數は全項を自己よりも大なる項と小なる項とに丁度二等分するところの中心値である。故に全項の中から隨意に取出した一項の値が中位數よりも大なる確率と小なる確率とは相等しいわけで、この意味から中位數を確率値といふこともある。中位數を基準にして測つた偏差の絶對値の總和は、他の如何なる値を基準として測つた場合のそれよりも小さい。これ散布度測定の一形態たる平均偏差を求むるに當り、特に中位數が推賞される所以である。（次章參照）。

（四）並數（モード）（Mode 最頻値）

一數列を構成する各項のうち、同じ値をもつものが繰返へし現はれる場合、最も屢々現はれる値をいふ。一六〇糎の身長を有する人が最も多ければ、この一六〇糎が並數である。度數系列については、度數の最も多い級が並數を含む級なることは明かであるから、その前後の二つの級の度數を考慮に入れて並數を決定する。前例を利用すれば、並數が第四級に在ることは明かであるから、その値は中位數と同樣に 16円＋x である。このxを次式で決定する。

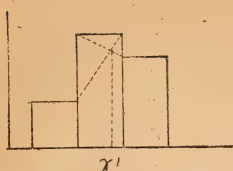
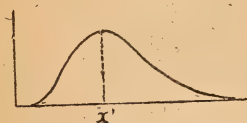
$$Mo = L_1 + \frac{f_1}{f_1 + f_2} C = 16 + \frac{58}{58 + 46} \times 4 = 18.23$$

Lは並數を含む級の最下限（この例では一六圓）f_1及びf_2はその前後の級の度數、Cは級の大さである。

近似的には並數は算術平均mと中位數Meとを利用して次式から求められる。

$$Mo = m - 3(m - Me)$$

この式は分布の對稱度を測るときに利用される（次章參照）。

中位數に於けると同樣、並數もグラフから求められる（第十二圖）。また解析的には、並數は曲線の極大値に對應するxの値であるから、$y = f(x)$の微係數を零ならしめるxの値x'を求めればよい。（第十三圖）

第十二圖

第十三圖

最も頻繁に現はれる値は、それが正常的な値たる證據と考へられる。即ち並數は正常値たる性

質を有する。そして我々は日常暗默裡に斯かる正常値によつて事物の代表値と見倣してゐるのである。それは必ずしも平均値として自覺されてゐないが、平均値の定義を一考すれば、並數がその一つたることは疑へない。出來合の洋服は平均的體格を標準にして作られるが、その場合の平均値は恐らくは日常の經驗から割出された並數であらう。即ち經濟生活に於ては並數は廣く利用されてゐるわけである。

中位數と同樣並數も亦極端な値によつて左右されない。極端な値は當然項數が少く、從つて並數の決定に全く影響を與へないからである。兩端が開いてゐる系列について計算されうることも中位數の場合と同じである。併し並數は全く偏差を度外視してゐるため、系列を敍述する値としては大なる欠陷があるのである。それは代表値としての價値しかないが、併しその限りでは大いに利用されて然るべきである。

第六章　平均値に關する若干の注意

平均の算式はかく幾種もあるが、問題はいかなる場合にいかなる算式を選ぶべきかである。算式が異れば答も異るから、算式の選定は正しく決定的意味をもつのである。ところが一般の統計學教科書には、これに關する説明が缺けてゐる。算式だけいかに覺えても、適當の利用法を知らなければ、役に立たない。若干の例によつて用法の一端を説明しよう。

平均壽命

生れた瞬間に死んで了ふ不幸な子もあり、百歳を超えてなほ矍鑠たる長命者もある。個々の人の壽命ははつきりしないが、併し平均的になら判る筈で、人生五十年といふ言葉は、昔の人の粗雜な經驗から歸納された大雜把な平均値を言つたものであらう。平均壽命は誰にも關係の深い問題で、また生活の合理化や醫學の進歩を最もよく反映する指標の一つでもあるから、之が正確な測定は統計家の重要課題の一つである。ところが一口に平均壽命といつても、實はいくつかの内容があるのであつて、一般にはそれらの區別が忘れられてゐるようである。

平均壽命には平均餘命と折半餘命との別がある。或る年齡の一群が全部死んで了ふまでに各人の生存した年數を加へ合せ、これを最初の人數で割つた値、卽ち一人當りの生存年數を平均餘命といふ。それは明かに算術平均である。注意すべきはこゝでいふ「或る年齡」とは必ずしも零歲だけでなく、いかなる年齡でもよいといふことである。現在二〇歲の者についても、五〇歲の者についても、計算できるのであつて、從つて單に平均餘命といつては無意味で、零歲の平均餘命とか、二〇歲の平均餘命とか、必ず年齡を限定しなければならないのである。生命表には各歲の平均餘命が揭げられてゐる。第六回生命表から若干を拔萃すれば次の通りである。

平均餘命

年齡	男	女
0	46.92	49.63
1	51.95	54.07
2	52.92	55.02
3	53.02	55.13
4	52.74	54.89
5	52.22	54.40
10	48.25	50.47
20	40.41	43.22
30	33.89	36.88
40	26.22	29.65
50	18.85	22.15
60	12.55	15.07
70	7.62	9.04
80	4.20	4.67
90	2.14	2.09
100	1.07	0.89

右の表から若干の重要な事實を讀みとることができる。第一は、女は男よりも概して長命で、ど

の年齢でも概して二三歳の相違があるといふことである。九〇歳頃には逆になるが、これは實はその頃まで長生きする男は甚だ少く（昭和五年には八〇歳の者は女四二、六九二に對して男は一五、四九八、九〇歳になると女三、二一八に對して男は僅か一、二四九、百歳以上の合計は女七四に對し男は三一人である）、謂はゞ篩ひ殘された壯健者ばかりになつて了ふからである。何れにしろ男女の壽命がかく異るのは、少くとも年齢の點では男女が異質物たる一つの證據で、兩者を一括した平均壽命の如きものは無意味にちかいのである。第二は、平均餘命の最も大きいのは、零歳ではなくて三歳前後だといふことである。これは乳幼兒は特に死亡の危險が多いからで、最危險期を脱したときが壽命の最も保證されるときなのである。第三に、現に例へば一〇歳の者が二〇歳に達しても、壽命は一〇年は短縮されないといふことである。右表では男は七・八四年、女は七・二五年だけしか短縮しない。これは矢張り、齢を重ねるに從つて弱者が淘汰されるからで、長生きすればするほど死ぬのを忘れると惡口されるのはこのためである。なほ壽命は時と共に改善される傾きがあり、第一回生命表では零歳の男は四二・八、女は四四・三で、男は四年强、女は五年强少なかつた。尤もこれは乳兒死亡率が改善されたからで、他の年齢ではそれほど大きな變化は起つてゐない。國による相違は極めて大きく、日本の零歳男子の四七年弱に對

して、瑞典のそれの如きは六三年強で、實に十六年以上の差がある。

平均壽命のもう一つの考へ方は折半餘命である。或る年齢の一群が次第に死んで減つてゆけば、數年後には丁度半分がなくなり、半分が生き殘る。それまでの年數を折半餘命といふのであつて、全員を死亡者と生存者とに二等分する點、即ち中位數に外ならない。生命表は人口十萬を單位として計算してゐるから、該表の各年齢の生存數を見て、丁度半分となつた箇所を探せばよい。第六回生命表の男子に於ては、約五二・七歳である。これは零歳男子が半減する迄の期間、即ち零歳男子の折半餘命で、平均餘命四六・九二に較べると六年近い差がある。平均餘命と同樣、われわれはいかなる年齢についてもこれを求めることができる。例へば二〇歳男子は、表の生存數は七一三一〇であるから、それが半分の三五六五五となる箇所は表によつて六四歳強なることが判る。即ち他の事情が大して變化しなければ今後四四年強にして半減すると豫定してよいわけで、これを平均壽命と見ることは極めて合理的なのである。平均餘命は前表に示した通り四〇・四一年で、約四年の差がある。平均餘命も折半餘命も共に合理的な平均壽命であるから、必要に應じてその何れかをとればよい。折半餘命の方が概して長いから、長生が希望なら、折半餘命を信じた方がよからう。但し高齢者に於ては平均餘命の方が長い。その說明は諸君にお委せし

よう。

電球の平均壽命を實驗するには一般に折半餘命卽ち中位數が用ひられる。算術平均だと、全部切れるまで待たねば計算できないが、中位數なら、半數が切れて了ふまで待てばよいから、時間的にも節約ができる。所謂「大量生産の統計的管理」の一例である。

二　平均結婚年齡

結婚といふ人生の大事を數字化して了ふのは聊か殺風景だが、人口統計の中には結婚に關する事項が極めて多い。月による結婚頻度、男子及び女子のそれぞれの結婚年齡、その差、經濟事情と頻度や年齡の關係、等々數へ出せばきりがない。いま平均結婚年齡について見るに、平均壽命と同樣、異る二つの內容がある。一つは算術平均により、他の一つはモードによるものである。

或る年に行はれた全結婚に於ける男の年齡を總計し、これを人數で割れば、算術平均による男子平均結婚年齡となる。女子についても同樣である。明治三二年には男二七・五八歲、女二二・九八歲であつたが、昭和一三年には前者は二九・九八歲、後者は二五・三四歲と何れも約二年半高まつてゐる。併しこれには再婚三婚も含まれてゐるから、別に初婚者だけの數字も出してゐる。

昭和一三年のそれは、男二八・三九歳、女二四・四一歳で、男は一歳半、女は一歳弱低い。男の再婚者には年寄の多い結果である。

然るに結婚年齢は女の場合には一五歳頃から始まつて急に高まり、二三歳前後で最頂點に達し、以後は次第に減少はしてゆくが、五〇歳や七〇歳になつてもまだ少しは結婚が行はれる。男子の場合には、二〇歳頃から始まつて急に高まり、二六歳頃頂點に達し、今後は女と同様ダラダラいつ迄も續いてゐる。即ち男女ともに結婚年齢曲線は甚だしい正の歪度を示してゐるのである。頂點の所在即ち男の二六歳と女の二三歳は、事實上最も多く結婚する年齢であるのに、前記の算術平均では、高齢者の年齢の影響で、この重大な事實が陰蔽されて了ふ。このことは總べての不平等分布についていへることで、例へば國民所得の増加は（人口が一定なら）一人當り所得即ち算術平均は増加せしめるが、もし不平等度が一層甚だしくなれば、モードは却つて低下するかも知れない。賃銀問題に於ても同様のことが起り得よう。

三　中間人口

年始と年末（又は次の年始）の人口が與へられたとき、中央の即ち六月末の人口を求める必要

が屢々起る。人口が算術級數的に增加(又は減少)してゐるならば、算術平均でよいが、日本のように幾何級數的に增加(又は減少)してゐるなら、幾何平均をとらねばならぬ。二つの國勢調査による確定人口に基き、その間の各年の人口を算出するにも同軌の方法が講ぜられる。前回の確定人口をA、五年後の今回のそれをBとすれば

$$A(1+r)^4=B$$

であるから、$1+r=\sqrt[4]{B/A}$としてr即ち平均增加率を算出することが出來る。これを利用すれば第二年、第三年、及び第四年のそれぞれの人口は$A(1+r)$, $A(1+r)^2$, $A(1+r)^3$となる。勿論增加が算術級數的か幾何級數的かは二つの數字(國勢調査)からだけでは判らない。少くも三つの數字が必要である。何れにしろ增加形態が不明な限り、何れの平均法がよいかは言へないわけである。

四　ケトレーの「平均人」

外國人を見慣れない人にとつては、人種の區別はつきかねるが、少し慣れると、一見して大體の見當はつくものである。ヨーロッパ人にとつては、日本人と中國人とは仲々見分けがつかない

らしいが、ベルリンや巴里の子供達は可成り目が肥えてゐて、私達も驚くほどよく當てたものである。一つの民族には固有の或る型があつて、それが構成單位たる個人に現はれるのであつて、それは肉體的な、また精神的な或る特徴となつてゐるのである。そこでわれわれが或る民族を考へるときには、かような特徴を想ひ浮べるのであつて、例へば英國人については、色白く背が高く、幾分痩せぎすで、むつつりはしてゐるが擧措の紳士的な人間を考へるし、獨逸人については、髪はブロンドで背はさして高からず、肥り氣味で、議論を好み、何となく垢拔けしない人間を想ふであらう。ヨーロッパ人は日本人とは黄色で貧弱で、眼鏡をかけカメラを携へ、深刻な顔つきをしながら、用もないのにニヤニヤ薄氣味わるい笑を洩す人間と見てゐるようである。

かような一般の見方には勿論誤解もあり偏見もあらうが、兎にかく凡ゆる民族に獨得の特徴のあることは事實である。併しこれを直觀的にいふだけでは餘りに非科學的である。近代統計學の開祖アドルフ・ケトレーは一國民の肉體的精神的道德的平均値によつてこれを數字的に確定せんとした。彼の有名な「平均人」(homme moyen) とは斯かる平均値によつて構成された一箇の假構物であるが、彼がかゝる假構物を必要と見たのは、畢竟社會の煩雜性はこの簡單な平均人のうちに總べて具現され、平均人の研究によつて社會の眞相を最もよく把捉できると信じたからで

人間は、社會の諸要素がその周りを動揺するところの一の平均である。……彼にとつては一切のことが社會に對して得られた平均的結果に適合して起るのである。もし吾々が一の社會物理學の基礎を幾らかでも確立せんとするならば、必ずこの種の人間を考案せねばならない」（平・山村共譯、人間に就いて、上卷三四頁）。

彼は肉體的標識（身長・體重等）は勿論、能力や犯罪性の如き精神的道德的諸標識も直接又は間接に計量しうるものとしたから、平均人が數字的に示されうることは當然として、問題は彼が平均人を以て眞善美の典型と見做したことである。曰く「一定の時代に於て平均人のあらゆる性質を自分の中に要約した一人の人は同時に偉大なるもの、美なるもの、善なるものの總べてを代表する」と（前掲書下卷二四二頁）。惟ふにこの考へ方は、正常的なるものと理想的なるものゝ混淆から生じた危險な結論の一典型たるものであらう。統計的平均値は集團に於ける大なるものの小なるもの、美なるものの醜なるもの、或ひは善なるものの惡なるものを一括した値である。それは大でもなく小でもなく、美でもなく醜でもなく、善でもなく惡でもないといふ意味に於て、確かに正常的ではあるが、併し決して理想的ではない。理想的なるものは、より美しく、より善

く、より眞なるものでなければならない。理想的賃銀は、勞働者にとつてはより高い賃銀であり、資本家にとつてはより低い賃銀であつて、平均賃銀は當事者何れの理想でもない。平均的體格を標準體格といふが、理想的體格は標準體格をよい方向に遙かに突破したものである。勿論凡ゆる意味に於て平均が理想的でないといふのではない。凹凸の甚だしい道路より平坦な道路がよく、明滅する燈りより均一な光の方がよい。但し一年を通じて平均温度を保つような土地が春夏秋冬の變化のある土地よりも快適かどうかは疑問であらう。併しケトレーのいふ平均とはかゝるものではなく、高い道路より少し低い道路がよく、明るい光より少し暗い光がよいといふのと同じであつて、殊に、異常な能力より平凡な能力がよく、異常な健康より中庸な健康がよいといふに至つては、正に凡俗の讃美に墮したものといへよう。統計學では、絶えず平均を以て他を判定する標準とするが、それはどこまでも形式上の標準以外の何物でもないことを銘記しなければならない。ケトレーの平均人は統計的思惟に於ける論理的飛躍の一つの見本として、幾多貴重な敎訓を含むものである。

註　異る二つの球體の平均半徑と平均重量をもつような球體はあり得ない。蓋し球の半徑をR

平均體積は

$$\frac{1}{2}\cdot\frac{3}{4}\pi(2^3+4^3)=\frac{3}{4}\pi 32$$

である。然るに二つの半徑の平均卽ち3を半徑とする球の體積は

$$\frac{3}{4}\pi 3^3=\frac{3}{4}\pi 27$$

であるから、平均半徑をもつ球の體積は、二つの體積の平均とは異る。このことからベルトランは、身長が平均身長に等しく、體重が平均體重に等しいといつた平均人は、もしありとすれば、一ケの怪物だと論じた。（エミール・ボレル著、矢野健太郎譯、偶然論、一七九頁參照）。併しこの批判にも誇張がある。六尺二十五貫の巨人と五尺十貫の小人を平均した五尺五寸十七貫半の人間、卽ちこの二人の平均體重と平均身長とを兼備した人間は充分ありうる。人間が球だとはケトレー自身いつてゐないのである。

五　客觀的平均値と主觀的平均値

平均値に關する最後の注意は、客觀的平均値と主觀的平均値の區別である。天體觀測なぞで判る通り、觀測には必ず誤差が伴ふ。山の高さは何度測つてもその都度多少の差は免れない。併し

觀測誤差は誤差法則に從ふから、數回の觀測結果を平均すれば誤差は相殺されて眞實の値が現はれる。即ち平均値は眞實値に外ならず、斯かる平均値を客觀的平均値といふのである。然るにわれわれが普通に言ふ平均値は多數の對象（例へば千人の身長）について求めるを原則とし、その場合には箇々の對象のそれぞれの値が眞實の値であつて、平均値は單なる代表値たるに過ぎない。それら箇々の値は觀測誤差と違つて誤差法則に從はないから、所謂正規曲線を描かず、從つて算式の如何によつて各種の平均値が現はれる。これを主觀的平均値といふ。

社會科學に於ける平均値は殆ど常にこの主觀的平均値である。故に平均値を以て眞實値と考へるのは明かに誤りである。尤も客觀的平均値も、それが眞實値たるがためには、誤差法則の作用する程度の多數の觀測値を必要とする。一課目について數回試驗を繰返へせば客觀的平均値が得られる筈であるが、實際にはこの回數は少いから、その平均點が直ちに眞實値だとは言へないであらう。併しこの種の平均値が、異る數課目の平均點と本質的に違ふことは明かである。

而もなほ主觀的平均値も、その方法さへ妥當なら、これを眞實値と見て大した誤りはないのである。殊に極めて安定した、即ち何度測つても略々類似の値を示すような平均値について然りである。後に述べるように、[illegible]

いふ理由でこれをも否定して云へば、統計的方法は全く無用の長物に過ぎない。但し社會的事實の統計的研究がこの點で大きな制限を受けてゐるといふ事實は、この方法に過度の信頼を寄せ易い最近の傾向に對して重大な警告として作用せねばならない。

第七章　分散度と非對稱度

一　分散度測定の必要さ

森に入つては森は見えないが森に入らなければ樹は見えない。平均値といふ要約的大觀的數字だけ振廻すのは、森だけ見て樹を無視することであり、天下國家を論じて身邊を忘れる類ひである。等しい平均値必ずしも等しい内容を意味しない。1と9の平均も、4と6の平均も、共に5である。だがこの場合5の意味は甚だちがふ。英語と數學の成績がA君は一點と九點、B君は七點と三點とすれば、平均點は等しいが、決して能力の等しいことを意味しない。通譯にはB君が採用されるだらうし、出納係にはA君の方が向くであらう。この場合には元々平均なぞもち出すのが間違ひで、必要なのは特定課目の成績點だけなのである。ところがどうしても平均値を問題にしなければならぬ場合がある。それは右の例でいへば、通譯を採用する場合、學校時代の二回の英語點數を調べたところ、A君は一點と九點、B君は三點と七點だつたとする。平均點は共に五點である。さてどちらを採用したらよいか。

これは上記の英語と數學の二課目の平均點とは甚だちがつた問題である。A君は大變英語が得意のようでもあるが、またそうでないようでもある。それに較べるとB君は可もなし不可もなしといつたところ。合計しても平均しても、問題は解決されない。採用する人にとつては兎に角考へさせられる問題であらう。

これは畢竟平均値をその內容と關聯させてどう解釋したらよいかといふことである。そしてこの種の問題は經濟問題でも絕えず起つてくる。例へば甲乙二村がある。村民の平均所得は共に等しいが、甲村は全部が中農なのに、乙村では大部分が貧農で、偶々一人の豪農がゐるため平均値は高くなつて、このため兩村の平均所得が等しいといふ場合はいくらでもある。この二村の所得狀態は全く異るに拘らず、平均値だけではこの點は不明である。要するに平均値は構成要素の多樣性を全く蔭蔽して了ふからで、問題はこの多樣性をいかにして表現しうるかといふことに歸着する。

構成要素のもつ多樣性は、平均値との關聯では、偏離或ひは偏差 deviation に外ならない。各要素の大さが何れも平均値と一致してゐるときは、各要素と平均値の間には開きがないわけで、これを偏差は零だといふ。普通の場合は、各項と平均値との間には種々の開きがあり、從つ

て種々の偏差がある。偏差は項の値と平均値との差であるから、前者が後者よりも大きいか小さいかによつて、正の偏差と負の偏差があるわけである。

各項の値が何れも相等しいとき、即ち何れも平均値に一致するときは、その數列は全く集中的だといえ、相互に大きな開きがあるときは、分散的だといふ。集中的か分散的かは、偏差の絶對値を合計して見れば判る筈で、集中度或ひは分散度は斯くて偏差によつて測定されるのである。そして集中とは分散の零の場合であるから、これらは一律に分散といい dispersion と名づけられる。

分散度は平均値の意味を確定する不可缺の數値である。或ひは平均値の代表性を明示する指標と言つてもよい。いくつかの數値を平均値といふ一箇の値に集約することは、畢竟全體を平均値を以て代表せしめることであるが、この場合、數列が集中的ならば、平均値は能く代表値たるの資格を有する。現實の値が何れもこれに近いからである、反之、分散的のときは、平均値は單なる計算上の値で、現實とは遊離した謂はゞ理論的假構物であり、從つて代表性は乏しいわけである。平均値の數學的性格は何れの場合でも同一であるが、これを現實理解の武器と考へるときは、代表性の大小が決定的重要さをもつのであつて、偏差の測定が統計學の重要課題たることも怪しむを要しないのである。そしてこれに關して幾多の方法が行はれてゐるがその最も代表的な

ものは標準偏差 (Standard deviation) である。

二　標準偏差

前述の通り偏差は平均値の抽象性を、換言すれば變量の分散度を判定せしめる標準である。既に述べた通り變量 x_1 x_2 x_3 …… x_n と任意の値Aとの偏差の總和、即ち

$$(x_1-A)+(x_2-A)+\cdots\cdots+(x_n-A)$$

は、Aが算術平均値に等しい場合には零である。また偏差の平方の總和、即ち

$$(x_1-A)^2+(x_2-A)^2+\cdots\cdots+(x_n-A)^2$$

は、Aが算術平均値に等しい場合に最小である。測定の基準としては斯かる最小値が最も妥當であつて、標準偏差 (σ) とはこの最小値の算術平均の平方根、即ち

$$\sigma=\sqrt{\frac{1}{n}\left[(x_1-A)^2+(x_2-A)^2+\cdots\cdots+(x_n-A)^2\right]}$$

のことである。算術平均値と各變量との偏差を d_1 d_2 …… と略記すれば

$$\sigma=\sqrt{\frac{1}{n}\,d_1^2+d_2^2+\cdots\cdots+d_n^2}=\sqrt{\frac{1}{n}\Sigma(d^2)}$$

と記せる。前例のA君については、$\sigma=\sqrt{\frac{1}{2}\left[(1-5)^2+(9-5)^2\right]}=4$　B君については、

$\sigma=\sqrt{\frac{1}{2}\left[(3-5)^2+(7-5)^2\right]}=2$　即ち一方は四點、他方は二點である。度數系列について

は、度數 f_1 f_2……f_n を考慮に入れて

$$\sigma=\sqrt{\frac{1}{N}(f_1 d_1^2+f_2 d_2^2+\cdots\cdots+f_n d_n^2)}=\sqrt{\frac{1}{N}\Sigma(fd^2)}$$

但し　$N=f_1+f_2+\cdots\cdots+f_n$

記の賃銀系列について計算すれば次の如くである。

中點	d	d²	f	fd²
¥6	−14.56	211.99	5	1059.95
10	−10.56	111.51	15	1672.65
14	− 6.56	43.03	46	1979.38
18	− 2.56	6.55	68	445.40
22	+ 1.44	2.07	58	120.06
26	+ 5.44	29·59	32	946.88
30	+ 9.44	89.11	22	1960.42
34	+13.44	180.63	10	1806.30
38	+17.44	304.15	2	608.30
42	+21.44	459·67	2	919.34
46	+25.44	647.19	0	0
50	+29.44	866.71	1	866.71
			261	¥12385.39

$\therefore\ \sigma=\sqrt{\frac{12385.39}{126}}=\sqrt{74.45}=6$圓88錢

二つの數列の分散度を比較するには、直接その標準偏差を比較してはならぬ。蓋し標準偏差は算術平均値を基準とした測定値であるから、算術平均値が等しくない限り、基準の大さが異るからである。斯かる場合には二つの標準偏差をそれぞれの算術平均値に對する比率に換算し、以て基準の大小を抽象する必要があるのである。即ち 二つの標準偏差をそれぞれσ'及びσ''二つの算術平均値をそれぞれm'及びm''とすれば

$$\frac{\sigma'}{m'} : \frac{\sigma''}{m''}$$

の形で比較しうるのである。斯かる比率を變化係數 (Coefficient of Variation) と名づける。この係數は標準偏差とちがつて比率即ち無名數なることを記憶せねばならぬ。一例を擧げよう。男工の平均日給八〇圓、女工のそれは五〇圓なるとき、前者のσは五圓、後者のそれは四圓だつたとする。σをそのまゝ比較すれば五對四で、男工に於ける分散は遙かに多いことになるが、變化係數はそれぞれ$\frac{5}{80}=6.25\%$及び$\frac{4}{50}=8\%$で、相對的には女工に於ける分散の方が大きいことが判るのである。

標準偏差は、その主たる用途は分散度の測定にあるが、同時に、一定の條件の下では、これと算術平均値との併用によつて數列の構造に關し貴重な知識を與へることが出來る。卽ちもし與へ

られた數列が、正常分布乃至それに近い分布をなすときは、算術平均値の上下にσに等しい幅をとれば、その範圍內に全項の約六八%が含まれるのである。身長の如きものは略々正常分布をなすから、例へば一萬人の身長平均が一六〇糎、σを五糎とすれば、約六千八百人は160±5即ち一五五—一六五糎の身長をもつと見てよいことになる。同様にして、平均値の上下に20をとれば全項の九五%が、30をとれば九九%即ち殆ど全部がその範圍內に含まれるのである。この理由は總べて誤差法則から導かれてゐるから、說明は後章に讓らねばならぬ。

三　其他の分散度測定法

數列の分散度を測定する手段として標準偏差の外に次の如きものがある。

(一) レンジ(Range, 極差又は偏差範圍ともいふ)。與へられた數列中の最大項と最小項の値の差である。株式市場なぞで株價の發表に、一々の株式につきその日に行はれた最高價格(高值)と最低價格(安值)を記すのはこの例である。動搖範圍を示すには最適であるが、少數の極端な値によつて決定される惧れがある。

(二) 平均偏差(Mean Deviation)。算術平均値又は中位數と各項の値との開きを合計し、項の

數で除したもの。合計の際、正負の記號を無視し、總べて正値として計算する。正負を算入すれば、基準を算術平均値にとれば偏差の合計は零になつて了ふ。平均偏差では偏差の大さだけを問題にしてゐるのである。平均偏差を算出するには、基準は算術平均値よりも寧ろ中位數の方がよい。蓋し中位數から測つた偏差の絶對値の總和は、他の如何なる點から測つたそれよりも小さいからである。

このことは解析的に次の如く證明できる。與へられた度數曲線を $y=\phi(x)$ とすれば、中位數はこの曲線の含む面積を二等分する點であるから

$$\int_{-a}^{0} \phi(x)\,dx = \int_{0}^{b} \phi(x)dx = \frac{1}{2}$$

中心をhに移して　$x=h$　とすれば

$$S(h) = \int_{-a}^{h} (h-x)\,\phi(x)\,dx + \int_{h}^{b} (x-h)\,\phi(x)\,dx$$

$$= \left[\int_{-a}^{0} + \int_{0}^{b}\right] (h-x)\,\phi(x)\,dx + \left[\int_{0}^{b} - \int_{0}^{h}\right] (x-h)\,\phi(x)\,dx$$

然るに

$$S(o)=\int_{-a}^{o}-(x)\phi(x)dx+\int_{o}^{b}x\phi(x)dx$$

$$\therefore S(x)'-S(o)=h\left[\int_{-a}^{o}\phi(x)dx-\int_{o}^{b}\phi(x)dx\right]+2\int_{o}^{b}(h-x)\phi(x)dx$$

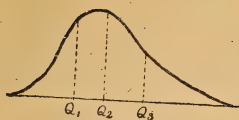
$$=2\int_{o}^{h}(h-x)\phi(x)dx$$

然るに$\phi(x)$は正の函數であるから、右の値は正である。即ち中位數を基準とした場合が最小である。

（三）四分位偏差（Quartile Deviation）。全度數を四等分する三つのx座標をQ_1 Q_2 Q_3とする（Q_2は中位數と一致する）。分布が集中的ならばQ_3-Q_1の幅は小で、分散的ならば大となるから、

$$\frac{Q_3-Q_1}{2}$$

によって分散度を測ることが出來る。これを四分位偏差といふ。（第十三圖）

Q_1　Q_2　Q_3

第十三圖

度數分布の解析に於て、集中傾向の程度即ち散布度を測定する必要あると同じく、分布の對稱の程度を測定する必要もある。全く對稱的な分布は理論的にのみ可能であつて、實際には多かれ少かれ非對稱的分布を示すからである。全く對稱性のない分布も勿論ある。所得稅納入者を所得額に應じて分類すれば次の第十四圖の如き所謂J字形をなすであらう。併し大部分の度數系列はどこかに中心を有し、歪んだ對稱性を示すものである。完全な對稱分布即ち正常分布を基準として、われわれは正と負の非對稱度を區別する。圖について知られたい。（第十五圖）非對稱度は歪度（Sk）とも呼ばれ、その測定には三つの方法がある。

（一）完全に對稱的な分布に於ては算術平均と並數は全く一致する。非對稱的分布に於ては兩者は相互に異り、且つ非對稱度が增大するに從つてその差は益々增大する。算術平均は極端な數値によつて最も强く影響されるから、非對稱度の增大と共に算術平均は益々並數（又は中位數）から遠ざかる。故に並數と算術平均との距離の大小は非對稱度の尺度たる事明かである。

然るに非對稱度は元來相互比較の爲に必要なのであるから、右の方法を以てしては次の二つの

困難に遭遇する。即ち、(1)比較さるゝ甲乙二系列の單位は必ずしも等しくない。(2)故に甲乙の非對稱度を比較する爲には、これを同一の基準に換算せねばならない。これには如上の距離を、それ〴〵の標準偏差に對する比率に書き改めればよいのである。即ち

$$S_k = \frac{M_a - M_o}{\sigma}$$

第十四圖

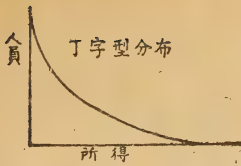
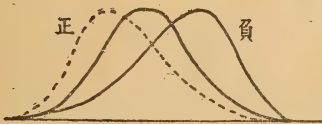
第十五圖

或ひは、既に説明した通り、略 $M_o = M_a - 3(M_a - M_e)$ であるから、これを右式に代入すれば

$$S_k = \frac{3(M_a - M_e)}{\sigma}$$

賃銀別職工表に就て見れば、M＝20圓56，Me＝18圓23 標準偏差は 6圓87 であつたから

$$S_k = \frac{3(20.56 - 18.23)}{6.87} = 1.02$$

この1・〇二は比率であつて、一圓〇二錢の意味でない事に注意されたい。上記二つの公式のうち、素より第一式、即ち並數を用ふるのがより適當である。これをピアソニアン非對稱係數

（二）非對稱度は四分位の位置からも測定される。對稱的分布に於ては第一及び第三四分位は中位數から等距離に在るが、非對稱分布に於ては然るを得ないから、この距離の大小によつて非對稱度を測定し得る理である。この距離を、四分位偏差に對する比率として示せばよいのであつて、即ち

$$S_k = \frac{(Q_3 - M_e) - (M_e - Q_1)}{Q.D}$$

前例に適用すれば

$$S_k = \frac{(24.50 - 19.82) - (19.82 - 16.00)}{4.25} = 0.2$$

この二つの方法の何れが勝れたりやは困難な問題である。第一の並數による方法は非對稱が小なる場合にも明瞭に指示すると言はれてゐるが、一般に並數の決定は困難な場合の少くない事を考へれば、中位數による代用法か又は四分位法を用ふる必要があらう。

（三）算術平均と各項との偏差の總和は常に零であるが、偏差の二乘又は三乘の總和は或る場合には零となり、或る場合には零とはならぬ。零となるのは完全な對稱分布に於てゞあり、非對稱分布に於ては零とはならず、且つ非對稱度が增大するに從つて零との距離は增大する。故にこの距離の程度から非對稱度を測定する事が出來る筈である。この方法に於ては偏差の三乘の總和

を用ひるのが最も適當である。蓋し二乘の場合には、負の偏差も正となり、從つて非對稱の方向が不明となるからである。なほ斯くて求めた三乘の總和の立方根を算出し、これを標準偏差に對する比率として示せば最も合理的なものとなる。即ち算式は

$$S_k = \frac{1}{\sigma}\sqrt[3]{\frac{\Sigma d^3}{N}}$$

となるべく、度數分布表に於ては

$$S_k = \frac{1}{\sigma}\sqrt[3]{\frac{\Sigma(fd^3)}{\Sigma f}}$$

となるのである。前例に適用すれば上の如く一・二一なる非對稱度を得る。

中點	f	d^3	fd^3
¥6	5	−3086.72	−15433.60
10	15	−1177.44	−17661.60
14	46	−282.08	−12975.68
18	68	−16.77	−1140.36
22	58	2.98	172.84
26	32	161.02	5152.64
30	22	841.20	18506.40
34	10	2432.64	24326.40
38	2	5304.37	10608.74
42	2	9648.00	19296.00
46	0	16464.77	——
50	1	25515.65	25515.65
	261		150789.91

$$S_k = \frac{1}{\sigma}\sqrt[3]{\frac{\Sigma(fd^3)}{\Sigma f}}$$

$$= \frac{1}{6.87}\sqrt[3]{\frac{150789.91}{261}} = 1.21$$

第八章　物價指數

一　物價指數の理念

物の價値は相對的で、常に他物との比較によつてのみ摑みうることは、最初に述べた通りである。物の經濟價値も亦同じで、交換經濟の下では、それと交換される他物の量で、貨幣經濟の下では、それと交換される貨幣の量で測られる。交換される貨幣量は即ちそのもの丶價値である。よつて一物の經濟價値はその價格によつて測られる。或ひは價格そのものと言つてもよい。ところが貨幣は、交換の媒介物で從つて價値の尺度であるから、當然自ら價値をもたねばならぬ。價値をもたぬものが價値の尺度となれないことは、長さをもたぬ物指が長さの尺度となれず、重さをもたぬ分銅が重さの尺度となれないのと同じである。然るに物指や分銅は不變なのに、貨幣價値は甚だ浮動的である。何となれば貨幣も亦他の商品と同樣に、その價値は貨幣の供給量と需要量によつて決定されるからである。このことは貨幣の最初の形態が實物貨幣即ち米とか貝とかであつたことを考へれば容易に納得できることである。貨幣價値のこの不確定性は價値尺度として

の貨幣の役割を甚だしく困難ならしめる。それは恰も、常に延びたり縮んだりする物指で長さを測らうとするやうなものである。物指とちがつて、貨幣價値に永久不變の一定の標準を求めても無駄である。併し物指が伸縮する場合、もしわれわれがその伸縮の程度を正確に即ち數字的に表現しうるならば、尺度として使用して差支へない。一尺の物指が半分に縮んだことが明かなら、それで測つた五尺は實は一丈であることが言明できよう。即ち貨幣については、永久不變の價値標準を言ふことはできなくても、或る時を基準としたその變化を測ることができさへすれば、この基準時を標準とした價値標準を、即ち購買力の變化を言ふことはできる。物價指數は元來はこの目的のために考案されたものである。

貨幣の購買力は一定單位の貨幣でどれだけの財が買へるかできまる。今まで百圓で米一升買へたのに、俄かに五合しか買へなくなつたとすれば、購買力は半減したわけである。ところがこのことは、別の言葉でいへば、今まで一升百圓だつた米が二百圓になつたといふことであるから、要するに價格が倍になれば、購買力は半減したことになるのである。よつてわれわれは價格の變化の逆數が購買力の變化を示すと言ふことができる。換言すれば、米の價格指數の逆數を見れば

らも變化する。年の豐凶による供給量の差、輸入量の變化、米に對する需要の變化等は、他の事情が等しくても、米價を變動させる。故に米價の變化だけから貨幣購買力のそれを知ることはできない。このことは如何なる商品についても言へることで、要するに特定の一商品の價格だけでは問題は解決されないのである。これを解決するには一商品の代りに、全商品をとればよい。何となれば貨幣は凡ゆる商品の購入に向けられるから、貨幣購買力の低下は一般に凡ゆる商品の價格を騰貴せしめ、貨幣購買力の騰貴は凡ゆる商品の價格を低下せしめる傾きがあるからである。箇々の商品の價格は、その商品の側の特殊事情に左右され、ときには貨幣購買力の低下したときに、或る商品の價格はさほど騰貴しないとか、場合によつては、逆に下落することもありうる。併し同じときに、他の或る商品の價格は、自己の特殊事情によつて、貨幣購買力の齎す騰貴より遙かに甚だしく騰貴するであらう。故にそれらを平均すれば正と負の異常變動は大體相殺され、貨幣購買力變動の影響を的確に知りうる筈である。かくてわれわれは全商品の價格指數を何等かの形で綜合し、單獨指數に代る綜合指數を作る必要があるのである。併し實際の問題としては、無限にちかい全商品をとり入れることはできないし、またその必要もないのである。

商品の中には重要商品と稱せられ、その取引のために多額の貨幣が動員されるものがある。石

炭、米、麥、生絲等々これであつて、一社會の貨幣の極めて多くの部分はこれが取引に充當されるから、これら商品だけを問題とすれば、貨幣購買力の變動は略々滿足に判るであらう。これから作られた綜合指數は多數商品の平均的價格即ち物價を示すから、物價指數と呼ばれる。

物價指數がその本來の目的たる貨幣價値或ひはその購買力の測定尺度たりうるかどうかについては、否定的見解も少くない。現實に極く限られた貨幣用途しか算入できない物價指數は、たかだか物價水準を示しうるのではあるまいか。貨幣の價値とか購買力といつたものは、畢竟實體のない假物ではあるまいか。これらの疑問は指數理論の根抵に觸れた重大問題であつて、到底こゝで取扱ふことはできない。併したとへ單に物價水準を測りうるに過ぎないとしても、それだけで猶ほ充分の使命は擔つてゐるのである。

かゝる商品の價格としては、自由價格制の下では卸賣價格またはそれに類する取引所相場をとるのが原則である。他の價格は純粹な經濟的原因以外の原因によつても動かされるから、甚だ區々で、標準となる價格を發見し難い。卸賣價格は自由競爭の結果略々一定し、所謂一物一價の原則が行はれるのである。取引所相場はその最も極端なもので、價格調査には最も都合がよい。併し統制經濟の下では主要價格は公定され、公價による卸賣物價指數は甚だ固定的になる。勿論貨

幣發行高もこれに應じて統制されゝば、物價水準は理論的にも安定するわけで、それが統制經濟の主要な目的である。併し現實には必ず闇價格が現はれ、公價による卸賣物價指數は現實と遊離し易い。今日物價指數が甚だ評判がわるいのは之がためであるが、この點については何れ後に言及するつもりである。こゝでは價格は適當に調査されたものとして、それをいかに指數化するかを考へよう。

二　指數の作製法

物價指數の第一の問題は綜合の方法で、大別して總和法と平均法の二つがある。各商品の價格を p', p'', p'''…………とし、基準時點のそれを p_0', p_0'', p_0'''…………比較時點のそれを p_1', p_1'' p_1''', ………… とすれば、比較時點の指數 I_{01} は

總和法では

$$I_{01}=\frac{p_1'+p_1''+p_1'''+\cdots\cdots}{p_0'+p_0''+p_0'''+\cdots\cdots}=\frac{\Sigma p_1}{\Sigma p_0}$$

平均法では、算術平均なら

$$I_{01}=\frac{1}{n}\left(\frac{p_1'}{p_0'}+\frac{p_1''}{p_0''}+\frac{p_1'''}{p_0'''}+\cdots\cdots+\frac{p_1^n}{p_0^n}\right)=\frac{1}{n}\Sigma\left(\frac{p_1}{p_0}\right)$$

幾何平均なら

$$I_{01}=\sqrt[n]{\frac{p_1'}{p_0'}\times\frac{p_1''}{p_0''}\times\cdots\cdots\times\frac{p_1^n}{p_0^n}}=\sqrt[n]{\Pi\frac{p_1}{p_0}}$$

簡單のために商品を米と石炭の二つとし、基準時の價格はそれぞれ五十圓（一升）と千圓（噸）、比較時點のそれはそれぞれ四十圓と千二百圓とすれば、總和法では $I_{01}=\frac{40+1200}{50+1000}=118$ 算術平均法 では $I_{01}=\frac{1}{2}\left(\frac{40}{50}+\frac{1200}{1000}\right)=100$ 幾何平均法では $I_{01}=\sqrt{\frac{40}{50}\times\frac{1200}{1000}}=98$ となる。總和法では一割八分の騰貴、算術平均法では騰落なく、幾何平均法では二分の下落である。併し總和法では單位はどうともとれる。米は石、石炭は粁瓦とすれば

$$I_{01}=\frac{400+1.20}{500+1.00}=80$$

となり、騰貴どころか、逆に二割の下落となる。然るに平均法ではこの例では結果は渝らない。

併しこれは米と石炭とが同一の經濟的重要さをもち、各々の取引に同額の貨幣が支出される場合に限る。米と石炭の代りに、米と鰹節をとれば、前者の二割の騰貴が後者の二割の下落によつて相殺されるとか、相殺されて餘りがあるとか考へることの如何に不合理かは明かである。總和法でもし適當に單位を選定すればこの不合理は避けられる筈だが、實際には困難である。惟ふにこゝでいふ重要さとは、その商品の取引のために動かされる金額であるから、それに從つて加重

すれば、問題は簡單に解決される。然るに取引金額は單位價格と取引量との相乘積であるから、いま取引量を q', q'', q'''……とすれば、取引金額はそれぞれ $p'q', p''q'', p'''q'''$……で示される。但しpに於けると同樣、qにも基準時點のそれ即ち q_0', q_0'', q_0'''……と比較時點のそれ即ち q_1', q_1'', q_1'''……がある。總和法では既に單位價格が與へられてゐるから、これに數量qを掛ければ取引金額となる。即ちウェートは取引數量で、q_0かq_1が用ひられる。平均法ではウェートは取引金額でなければならぬから、p_0q_0、p_0q_1、p_1q_0、p_1q_1 の四つの場合が考えられる。併し實際には p_0q_0 か p_1q_1 が用ひられる。即ち加重總和法は

$$I_{01} = \frac{p_1'q_0' + p_1''q_0'' + p_1'''q_0''' + \cdots\cdots}{p_0'q_0' + p_0''q_0'' + p_0'''q_0''' + \cdots\cdots} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \cdots\cdots \text{Laspeyres式}$$

$$I_{01} = \frac{p_1'q_1' + p_1''q_1'' + p_1'''q_1''' + \cdots\cdots}{p_0'q_1' + p_0''q_1'' + p_0'''q_1''' + \cdots\cdots} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \cdots\cdots \text{Paasche式}$$

の二つがある。前者をライパイレス式、後者をパーシェ式といふ。これに對して、加重算術平均法では

$$I_{01} = \frac{\Sigma\left(\frac{p_1}{p_0} \times p_0 q_0\right)}{\Sigma p_0 q_0} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \cdots\cdots \text{Laspeyres式}$$

$$I_{01}=\frac{\Sigma\left(\frac{p_1}{p_0}\times p_1q_1\right)}{\Sigma p_1q_1}$$

$$I_{01}=\frac{\Sigma\left(\frac{p_1}{p_0}\times p_0q_1\right)}{\Sigma p_0q_1}=\frac{\Sigma p_1q_1}{\Sigma p_0q_1}$$ ……Paasche式

$$I_{01}=\frac{\Sigma\left(\frac{p_1}{p_0}\times p_1q_0\right)}{\Sigma p_1q_0}$$

の四つがあるが、うち二つは加重總和法と同じラスパイレス式とパーシェ式になる。

加重幾何平均法には

$$I_{01}=\sqrt[\Sigma p_0q_0]{\Pi\left(\frac{p_1}{p_0}\right)^{p_0q_0}}$$

$$I_{01}=\sqrt[\Sigma p_1q_1]{\Pi\left(\frac{p_1}{p_0}\right)^{p_1q_1}}$$

等々の算式か得られる。平均法で實際に最も廣く用ひられるウェートはp_0q_0即ち基準時點の取引金額である。

以上の如く指數算式の種類は多いが、ではどれがよいか。結局ここで問題となるのはラスパイ

レス式とパーシェ式であるが、その差は前者はウェートとして $p_0 q_0$ を、後者は $p_1 q_1$ を用ひてゐることである。それが如何なる結果を與へるかは指數論の大きな問題であつて、こゝにその詳細を說くことはできないが、結論的に言へば、ライパイレス式は實際よりも低い結果を示し、パーシェ式は實際よりも高い結果を示す傾きがあるのである。故に最も適當な指數式は右の二つの算式の平均をとるべきで、その際算術平均をとるか幾何平均をとるかによつて次の二つの式が得られる。

$$I_{01} = \frac{1}{2}\left(\frac{p_1 q_0}{p_0 q_0} + \frac{p_1 q_1}{p_0 q_1}\right)$$

$$I_{01} = \sqrt{\frac{q_1 q_0}{p_0 q_0} \times \frac{p_1 q_1}{p_0 q_1}}$$

後者は「フィッシャーの理想式」と稱せられるもので、最も廣く支持されてゐる。

最古の卸賣物價指數は一八五九年ニューマーチ（W. Newmarch）の作製にかかる。但しこれはいくつかの價格指數を列記したゞけで、總平均は計算されなかつた。この指數は一八六四　經濟雜誌エコノミスト（The Economist）に繼承されたが、一八六九年から總指數が作られるに至つた。最初の品目は二十二、方法は算術平均法であつた。一八九九年から每月發表に改められ、

その後品目も増加し、一九二八年末からは幾何平均法によることゝなつた。この指數と共に古い歷史を誇るものにステーティスト指數がある。サウエルベック（A. Sauerbeck）が一八八六年に創始したもので、一九一二年末から經濟雜誌ステーティスト（The Statist）に引繼がれた。品目は約五十、方法は算術平均法である。

米國の最古の指數はブラッドストリート（The Bradstreets）のそれである。一一一品目の各一ポンド價格を合計したもので、指數の名稱は不適當であるが、基準さへきめれば總和法によつて指數化は容易である。この社は一九三三年ダン社（Dun & Co.）と合併し、ダン・アンド・ブラッドストリート指數として今日に及んでゐる。

我國では明治二十年代から試作され出したが、日銀が今日の東京卸賣物價指數を作り始めたのは明治三十三年（一九〇〇年）からである。品目は五十六、方法は單純算術平均法であつたが、昭和十一年十一月から品目は百十、算式も加重算術平均法に改められた。品目としては、原則として、昭和七年から九年までの品目別內地生產價額及び同輸移入價額合計の年平均から、取引に上らぬ自家消費推定額を控除したものが千萬圓を超える商品に限定し（若干の例外もある）、ウェートとしては大體基準時千萬圓を以て一としてある。即ちそれはラスパイレス式といつてよ

い。日銀指數の外には商工省の全國卸賣物價指數、東洋經濟新報社及びダイヤモンド社のそれぞれの東京卸賣物價指數が有名である。殊に後者は加重幾何法による唯一の指數として注目される。

三　現在の指數問題（實效價格とパリティー計算）

物價指數は價格といふ經濟の中心問題に關するから、經濟學と特別の關係がある。ジェヴォンスがこを採り上げて以來指數理論の進歩に大きな貢献をしたのはエッジウァース、フィッシャー、ミッチェルの如き經濟學者であつた。そして最近に至つて指數理論は計量經濟學（エコノメトリックス）によつて飛躍的進步を遂げんとする情勢にある。舊時の理論では取引される各商品の價格と數量とは何れも獨立變數と見、綜合的な物價指數をこの二組の變數の或る一定の函數として表現しようとしたが、その函數の形は形式的に決定され、經濟理論的裏づけを缺いてゐた。ハーバラーに始まる最近の指數學者は價格と數量間に存する經濟理論的關聯を前提として、新らしい指數理論を展開するに至つた。フリッシュは舊理論を原子論的、新理論を函數論的と名づけた。この新らしい傾向はアレン、ステーレ、ワルド及び特にフリッシュによつて代表されてゐる。元々、指數の中心課

題は、物價が變動したとき購入者が以前と同じ滿足をうるために支出金額をどれだけ增加又は減少したらよいかといふことである。併しこのためには消費習慣を不變と假定せねばならないが、その結果、時間的場所的比較は現實から著しく遊離せざるを得ない。この困難はフリッシュが彈力性概念から導いた「限界效用の新測定法」によつて克服されんとしてゐる。これら理論的問題はこゝでは論述の範圍を超えてゐる。特志の讀者は森田優三氏の「物價變動の測定」について知られたい。

物價指數は物價の變動を測定するのが直接の任務であるから、それが最も要求されるのは言ふまでもなく物價變動の最も激しい時代である。この意味で、今日ほどこれが切實に要請される時代はない。ところが同時に今日ほどこれが作製の困難な時代もないのであつて、多少とも信用しうる物價指數は皆無といふのが眞相である。上述の通り、物價指數は市場價格を前提とする。市場價格とは需要と供給の關係から定まる自然的な價格で、從つてそれは自由經濟の下にのみ可能である。ジェヴォンスはこれを一物一價の原則と名づけた。ところが統制經濟の下では價格は公定され、需要と供給はいかに變化しても、價格自體は容易に變へられない。故にもし公定價格が嚴守されゝば、物價指數は殆ど固定し、指數の使命は沒却される。もし嚴守されなければ、闇價

格が行はれ、一物に無數の價格が成立するから、調査自體が不可能になる。勿論闇價格もいつしか相場をもつことになり、或る程度の標準はでき上つて了ふから、その曉には指數化も全く不可能とは言へない。併しそれは違法を承認することゝなり、公然發表することはできまい。即ち發表そのものが闇の性質を帶びることになるのである。

日本の現在では公定價格と闇價格とは同時に存在し、價格體系を一層混亂させてゐる。これに對處するために發明されたのが所謂實效價格の計算である。一商品の公定と闇の價格と、それぞれの價格で取引される數量とが推算されゝば、單位當りの平均價格が計算できる。家庭で米を配給で二〇キロ、闇で一〇キロ買へば、前者は例へば一キロ十五圓、後者は百圓として、支出額は20×15＋10×100＝1300圓 即ち一キロ當り四十三圓强となる。次表は總理廳統計局調「消費者價格調査」による白米の公定價格、闇價格及び實效價格である。勿論公價、闇價及び實效價格の動きは適當の時點を基準として指數化することができる。時事新報社刊「統計に見る經濟危機の實相」六一七頁を參照されたい。實效價格は闇依存度いかんによつて大差があるから、諸君自ら家計簿によつて算出されたい。

重要商品全部について實效價格が求められゝば、それらを綜合して實效物價指數を作りうるわ

期間	公定		闇		實效價格
	價格	配給量	價格	購入量	
21年 7月上旬	2.00	0.530	51.65	2.45	42.89
11月上旬	3.62	11.830	37.10	1.15	6.61
22年 6月下旬	3.60	3.310	127.66	3.89	50.11

けである。併し闇價格の不統一性と、公價と闇價で購ふ數量の割合の不統一性とを考へれば、それが殆ど不可能なことが判る。

右に次で最近の指數計算を困難ならしめてゐるものは品質の變化といふことである。名稱は變らぬが品質は一變して了つた商品のいかに多いかは、誰しも日常經驗してゐる。品質が半減すれば價格も半減されて然るべきで、價格が變らなければ實は二倍に騰貴したことになるのである。品質が改良されたときは勿論その逆でなければならない。品質の變化は原料よりも半製品に於て、半製品よりも完成品に於て、これを測定すること益々困難である。石炭ならカロリー計算によつて科學的に行へよう。併しラジオ聽取器の音響や美觀や耐久力の改良または改惡はいかにして正確に測りうるか。これが計算され適當に算入されない限り、價格に關する總べての指數は甚だ無力と言はねばならない。そして、品質のこの問題が生産量指數にも決定的影響を與へることは後に述べるであらう。併し右に述べたように、品質の變化は完成品に於けるほど測定困難であるから、主要原料品に重點を置く物價指數よりも、日常品を中心

とする生計費指數に於てより問題とされねばならぬ。

パリティー計算

公正な價格は生産費によつて決定されねばならぬ。故に公定價格は原則として生産費によるべきであり、事實久しきに亘つて實行された。併し基礎資料の缺如や企業者間の生産費の著しい相違なぞによつてこれが不適當となつたとき、パリティー計算（Parity Calculation）と稱する別の方法が案出された。我國では米價は物價體系の中心であるから、その合理的決定は最も重要な問題の一つである。食糧管理法の規定では生産費主義に基くことになつてゐたが、昭和二十一年九月の米價改訂に於てはパリティー方式が採用された。その後の蠶絲及び木材の公價決定も亦これによつてゐる。パリティー計算とは或る物の價格(A)を決定するに當つて、その物の基準價格(A_0)へ、これが生産に關係ある諸項目の價格變動率$\left(\frac{B}{B_0}\right)$を乘ずる方法である。即ち

$$A = A_0 \times \frac{B}{B_0}$$

となる。上記の米價決定に於ては昭和九年から十一年までの平均價格（二八圓四八錢）へ、農業用品十一品目と家計用品二十一品目の平均騰貴率一、九五九%を乘じた $¥28.48 \times \frac{1,959}{100} = ¥558.00$ を算出し、これを五五〇圓として發表したのである。家計用品の比重が大きいのは、米

作に於ける生産費の中心が勞働にあるからで、勿論各品目のウェートは農家經濟調査によつて定められた。算式はフィッシャー理想式である。生産費主義によれば生産費の昂騰は直ちに公價の引上げを必要とし、公價の引上げは更に生産費の騰貴を促すといふ惡循環を招く惧れがある。今日の如き異常なインフレ時代にはこの點パリティー計算は大きな强味をもつものといへる。併しこれはどこ迄も一時的便法で、眞の公正價格が生産費主義によらねばならぬといふ原則は渝らないのである。

第九章　生計費指數

一　家計調査と生計費指數

働かざる者は喰ふべからずといふ。併し働けば食へるかといへば、問題は勞働報酬で生計が購へるかどうかで決まることである。ところが激化するインフレーションの下では、勞働報酬は物價の昂騰に及ばず、從つて購へる生計は次第に貧弱になる。即ち生活水準は低下せざるを得ない。敗戰國民が急に以前よりも豊かな生活を望むのは無理だが、最低生活さへ保證されないとすれば、問題は深刻である。

生計費は物價とは一致しない。後者は理論的には全商品價格の、實際には重要商品價格を卸値で捉へて平均したものである。ところが生計費は一家の日常生活に必要な支出で、小賣値で買ふ上に、重要商品たる原料品なぞは殆ど項目に入らないし、たとへ入つてもその割合は極く少い。その上物價指數では全く問題外に置かれた家賃・醫療費・交通費・税金・敎育費といつた項目が重きを占める。故に生計費の變化は物價指數では滿足には測れないのであつて、別に生計費指數

なるものが作成されなければならない。これによつてわれわれは、一定の生活を營むために必要な費用がいかに變化したかを知ることができ、これから多くの政策を導くことができるのである。例へば合理的な賃銀はこの指數に應じて決定されねばならず、生活扶助・家庭手當・扶養家族の課税控除等々の社會政策的措置も總べてこの指數と睨み合せて考慮されなければならない。社會的困窮が加重し、國民の大部分が貧民階級に轉落した今日、この指數の重要さはこれを如何に高く評價しても足りないであろう。併し平時ならばその作製は格別困難ではないが、今日では幾多の致命的障碍が横はつてゐて、正確な指數は殆ど算出し難い狀態である。原理的には生計費指數の問題は簡單である。即ち多數の家族の支出狀態を平均的に算出し、支出項目とウエートを決定し、これに基いて價格變動を求めればよいのである。第一の手續きを家計調査といふ。

今から百二十年前、フランスの著名な社會學者ル・プレイ（Le Play）は歐洲諸國をはじめ世界各地に亘つて、集約的家族調査を開始した。國民の大部分を占める勞働階級から代表的家族を選出し、彼等の生態を詳細に實地調査することによつて各國の安寧度を求めんとするのが目的で調査員は數日乃至數週間その家族と起居を共にし、その間絶えず質問や觀察を續けて廣範圍の事項を調査したのである。收入及び支出狀態はその一項目に過ぎないが、詳細を極めた。研究の結

頁	行	誤	正
[illegible]	4	20	2σ
〃	5	30	3σ
〃	13	合計のし、	合計し、
109	6 7	$I_{01}=\frac{1}{2}\left(\frac{p_1q_0}{p_0q_0}+\frac{p_1q_1}{p_0q_1}\right)$ $I_{01}=\sqrt{\frac{p_1q_0}{p_0q_0}\times\frac{p_1q_1}{p_0q_1}}$	$I_{01}=\frac{1}{2}\left(\frac{\Sigma p_1q_0}{\Sigma p_0q_0}+\frac{\Sigma p_1q_1}{\Sigma p_0q_1}\right)$ $I_{01}=\sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0}\times\frac{\Sigma p_1q_1}{\Sigma p_0q_1}}$
162	7	圖の點線	圖の實線
163	7 9 10	$\frac{es}{es}$, $\frac{es}{eb}$	$\frac{\partial s}{\partial s}$, $\frac{\partial s}{\partial b}$
169	11	$y=ab^{Cx}$	$y=ab^{c^x}$
〃	12	$\log a+cx\log b$	$\log a+c^x\log b$
173	10	$\frac{X_1}{x}$, $\frac{X_2}{x}$, $\frac{X_{12}}{x}$	$\frac{X_1}{\bar{x}}$, $\frac{X_2}{\bar{x}}$, $\frac{X_{12}}{\bar{x}}$
243	13	$\frac{\sigma'}{\sigma''m'}$, $\sigma\frac{\sigma'}{n''}$	$\frac{\sigma'}{\sqrt{m'}}$, $\frac{\sigma''}{\sqrt{n}}$

頁	行	誤	正
1	6	無駄	無闇
7	9	美しさ	好しさ
54	10	po,	p_0,
64	7	$n\times\Sigma M=(x)^\circ$	$n\times M=\Sigma(x)$
65	4	$\Sigma(x-M)_2$	$\Sigma(x-M)$
67	7	$^n\sqrt{\pi(x)}$	$\sqrt[n]{\Pi(x)}$
〃	9	$\Sigma(f)\sqrt{\Pi(xf)}$	$\sqrt[\Sigma(f)]{\Pi(x^f)}$
71	5	$\int_2^{x'} f(x)dx=$	$\int_\alpha^{x'} f(x)dx=$
73	4	$m-3(m-Mo)$	$m-3(m-Me)$
90	4	だとしえ,	だとしさ,
91	10	$\cdots+(x^n+A)^2]$	$\cdots+(x_n-A)^2]$
92	表ノ下	$\sigma=\sqrt{\frac{12385.39}{126}}=\sqrt{74.45}$	$\sigma=\sqrt{\frac{12385.39}{261}}=\sqrt{47.45}$

果は續々「社會經濟協會」(Société d'Economie Sociale)の機關紙「二つの世界の勞働者」(Les Ouvriers des Deux Mondes)に發表されたが、元々箇別研究であつたから、單行論文の形式がとられた。ところがル・プレイの門下生にエルンスト・エンゲル(Ernst Engel)といふ俊才が現はれた。エンゲルはプロシャ人で、最初は技術家であつたが、後に社會問題に沒頭し、ル・プレイの指導を受けた。後年はプロシャ統計局長に就任し、所謂エンゲル法則の發見者として不朽の名を殘すに至つた。

エンゲルは家族の生活に於ける經濟的要素を特に重視し、レ・プレイに於ては單に一項目に過ぎなかつた收支關係を出來うる限り明瞭に把握せんと志した。勞働家族といつても相互に收支は等しくないから、箇別調査では一般的結論は求め難い。エンゲルはこゝに着目し、ベルギーの多數勞働家族について統計的調査を重ね、一八九五年「家計調査より求めたるベルギー勞働家族の生計費の變遷」を著した。(本書は森戸辰男氏によつて邦譯され、統計學古典選集に收められてゐる)。エンゲル法則とは、生計費の中で食費の占める割合が多いほど、其の家族は貧しいといふことで、この割合をエンゲル函數といふ。また彼はケット(Quet)と稱する消費單位を決定することによつて、異る家族構成の生計必要額を計算しうる途を拓いた。家計調査では所謂標準家

族を選定するとはいへ、必ずしも人數・年齢・體性の全く等しい家族のみを集めるわけには行かない。而もこれらの點が異れば、生計狀態は自ら異るわけで、相互の比較は不可能になる。そこでエンゲルは零歳の乳兒の消費量を一ケットと命名し、年齢が一歳加はる毎に、男子は滿二十五歳まで、女子は滿二十歳まで、年々〇・一ケット增加するものとした。卽ち成年男子は三・五ケット、成年女子は三ケットとなる。父母と一〇歳と八歳と六歳と四歳の四兒から成る家族は

$$3.5+3.0+2.0+1.8+1.6+1.4=13.3$$

の消費單位に換算されよう。ケットの名稱は近代統計學の父ケトレー（Quetelet）を記念するためである。エンゲルの業績は明治十年頃既に杉亨二、吳文聰なぞの先輩によつて我國に傳へられたが、實際に我國に始めて應用されたのは、恐らく高野岩三郎氏による「東京に於ける二十職工家計調査」であらう。大規模には內閣統計局が大正十五年に一年間に亘つて十九府縣道の給料生活者（官公吏、銀行會社員、敎員、巡査）、工場勞働者、鑛山勞働者、交通勞働者、日傭勞働者及び農業勞働者計七八五六世帶について實施した。昭和六年六月內閣統計局はこの調査の結果から全國勞働者及び給料生活者生計費指數ウェートを發表し、朝日新聞社はこれに基いて昭和六年十月以降每月全國生計費指數を作製發表するに至つた。一方內閣統計局は米穀統制の基本資料を

求める目的から昭和六年以降全國主要十都市について家計調査を續行することゝなり、大正十五年の調査よりも世帯を減らして、給料生活者五百乃至六百世帯、勞働者世帯千百乃至千二百を定した。調査方法は家計簿記入式で、選擇された世帯は一年間收支の詳細を記入せねばならぬ。世帯選擇の標準は（一）平均月收五十圓以上百圓未滿なること、（二）世帯主の勤勞所得を主たる收入とし、營業をもたぬこと、（三）無償で他から食糧その他の生活必需品の支給を受けぬこと、（四）世帯員は合計二人乃至七人で、同居人や家事使用人のないこと等を條件とした。實物收入は常にその評價額を記入させる。

內閣統計局はこの資料を利用して新たにウェートを決定し、昭和十二年七月、即ち日華事變勃發と共に、自ら生計費指數を作製するに至つた。調査は二十四都市について行はれたが、これから全國指數を作るためには都市の大小をウェートとせねばならぬ。所謂「人口ウェート」これであるが、注意すべきはこのウェートは當該都市人口そのものではなく、その都市の代表する人口だといふことである。卽ち全國都市一二五市の人口を北海道、東北、關東等十區に分ち、各區の都市人口をその區にある生計費指數調査都市に比例せしめたのであつて、表示すれば次の如くである。

入口ウェート

札幌市	4	大阪市	17
仙臺市	3	神戸市	5
山形市	1	鳥取市	1
郡山市	1	岡山市	2
前橋市	1	廣島市	4
東京市	34	德島市	2
横濱市	4	今治市	1
新潟市	1	八幡市	4
金澤市	2	長崎市	4
松本市	2	熊本市	3
濱松市	1	延岡市	1
名古屋市	9		
京都市	6	計	113

ウェートの單位は大體人口二十萬人である。京都市や神戸市のウェートが小さいのは附近に調査都市が多いからで、反對に、名古屋市や八幡市のそれの大きいのは附近に調査都市がなく、從つて多くの都市を代表せねばならぬからである。なほ調査が都市に限定されたのは、この指數が勞働者及び給料生活者といふ都會性階層に關するものだからである。

二　實質賃銀とスライド制

賃銀その他の所得の價値は支出と相對的である。賃銀は元々一家の生活を維持するためのもので、事實それ以外に使へるほど高額なものは稀である。して見れば賃銀の價値は、生活に必要な支出いかんによつて定まることで、後者が增大すれば前者は低下し、後者が減少すれば前者は增

加する。單に月給五千圓といつただけでは多いのか少いのか見當もつかない。故にこの名目賃銀を生計費變動と比較する必要がある。これを實質賃銀といふ。方式は簡單で、名目賃銀を生計費指數で割ればよい、卽ち

$$實質賃銀 = \frac{名目賃銀}{生計費指數}$$

である。戰終時の名目賃銀を一〇〇圓、現在のそれを二〇〇〇圓とし、當時の生計費指數は八〇、現在のそれは三〇〇〇とすれば、この時期のそれぞれの實質賃銀は

$$\frac{100圓}{80} = 125 \qquad \frac{2,000圓}{3000} = 66.7$$

で名目上の二十倍の引上げに拘らず、實質的には $\frac{66.7}{125} = 53.4$ 卽ち約四割七分低落したことになる。別の言葉でいへば、一〇〇圓が五三圓に引下げられたことゝ同じなのである。賃銀問題は斯くて常に實質賃銀の問題として取扱はれねばならない。そのためには正しい生計費指數が不可欠前提である。

さて終戰後の急激な物價騰貴は忽ち勞働階級を脅かし、賃上げ爭議は激增の一路を辿つてゐる。物價が上昇を續けてゐる間は、一度で引上げられた賃銀も忽ち生活費に追越され、絕えず爭議を繰返へさねばならぬ。この不便を避けるには賃銀を生計費の昻騰に自動的に追隨せしめる方

式が必要となるのであつて、所謂スライデング・システム（Sliding System）これである。この制度は別にスライディング・スケール制、賃銀滑準法または傾斜賃銀制とも呼ばれ、更に略して單にスライド制とも稱される。例の東芝爭議で勞働者側から要求されたが、採用されたのは電產爭議に於てゞある。今日では廣く行はれてゐるが、併し內容は區々で、杜撰なものが少くない。もし正確な生計費指數が得られゝば、基準賃銀にこの指數を乘ずることによつて、實質賃銀は常に同一に保たれる。卽ち、基準時點の名目賃銀をN_0、生計費指數をI_0とすれば、そのときの實質賃銀R_0は

$$R_0=\frac{N_0}{I_0}$$

である。同樣にして比較時點の實質賃銀R_1は

$$R_1=\frac{N_1}{I_1}$$

である（N_1は比較時點の名目賃銀、I_1はそのときの生計費指數）。問題はN_1をいかに決定すれば$R_0=R_1$となるかであるから

$$\frac{N_0}{I_0}=\frac{N_1}{I_1}$$

ならしめるようN_1を決定すればよい。即ち

$$N_1 = \frac{N_0 \times I_1}{I_0} - N_0 \times \frac{I_1}{{}_{01}} = N_0 \times I_{01}$$

である。I_{01}とは基準時を出發點として計つた生計費の變化率である。故に先月の賃銀は千五百圓、先月の生計費指數は三〇〇だつたものが、今月は後者が三二〇に上昇したとすれば、賃銀は$¥1500 \times \frac{320}{300} = ¥1600$とされて始めて同一の實質賃銀となるわけである。理論的には斯く簡單であるが、現在スライディング・システムが疑問視される主なる理由は次の如きものである。

(一) この制度は基準時の實質賃銀を保持する爲のもので、少しもその增加を目指してゐない。故に基準時のそれが不當に低かつたとすれば、この制度によつてその不當さがいつ迄も是正されない。

(二) 最近の賃銀は基本給よりも手當其他の形をとる部分が少くない。故に基本給だけにこの制度が適用されゝば、實質賃銀は次第に低下する。併し實物給與の分にまでこれを適用することは不可能にちかい。

(三) 生計費指數は作製から發表までに可成りの日時を要するから、その間の物價騰貴はとり入れることが出來ない。即ちそれだけ賃銀は後ろにあるわけである。この理由によつて、權威ある生

計費指數の發表を待たず、勞組自身拙速的にこれを作製する傾きがあるが、この場合には指數を不當に引上げるような計算も行はれないではない。反對に會社側にやらせれば、勿論なるべく低い結果の出るような方法が講ぜられよう。かくては指數の妥當性そのものが爭議の種とならざるを得ない。而も終戰後は官廳の指數も決して信賴しうるものではない。本來の意味に於ける家計調査は久しく中絶し、極めて小範圍の調査が行はれて來たに過ぎない。即ち安じて基準となしうる指數がない以上、いつ迄も水掛け論が續くであらう。

（四）　生計費指數の上昇は必ずしも生活內容の低下を正直に示さない。質の低下が充分反映されないからである。月八〇圓で濟んだ食費が八〇〇圓に上つたとすれば十倍に當るが、實はこの八〇〇圓で購ふ食物內容は八〇圓時代のそれとは比較にならぬお粗末なものである。この質の低下を以前の半分とすれば、實際の値上りは二十倍でなければならない。生計費指數にはこれは全く現はれないから、スライド制によればそれだけ勞働者側は損をすることになる。いまこれを書いてゐる最中、四貫俵木炭が配給された。殆ど粉ばかりでは半分も使へそうには見えない。値段は九四圓五〇錢といふから、二圓五〇錢時代の三十八倍である。指數とすればそうだが、半分しか使へないとすれば、二貫目で九四圓五〇錢、即ち七十六倍が本當である。品物の全部が全部これほ

どではあるまいが、とにかく指數に比例した賃銀では生活は可成り低下することは爭へない。

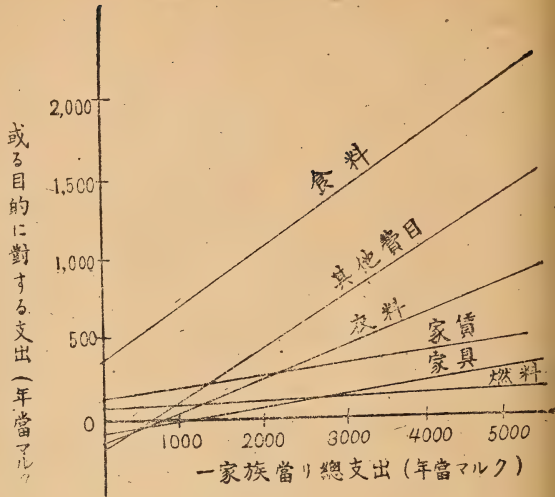
かくて最近の勞働爭議は單純なスライド制より更に一歩を進めて、寧ろ公正な賃銀形態の確立を目指してゐるものが多い。生計費指數によるスライディング制と並んで、例へびカロリー計算方式が要求されてゐるのはその現れである。後者は一定のカロリー（例へば二千四百カロリー）を確保しうる賃銀を要求するもので、一見甚だ科學的に思はれる。所得の七割前後が食費に充當される今日、賃銀の基準を食物に求めることは自然であるが、一定カロリーを攝取するに必要な金額は極度に區々で、例へば百カロリーは配給米なら十錢札で足りるが、闇の高級魚では數十圓も必要である。價格と必然的關係のないカロリーを賃銀と結びつけるのは根本的に無理があるのであつて、本來なら一顧の價値もない筈だが、それが盛んに論議されるのは、生計費によるスライディング制に重大な欠陷がある證據でもあり、同時に勞働爭議が賃銀なるものの本質にまで及んできた苦悶の現はれでもある。戰鬪的な勞働組合ほどカロリー計算を主張してゐることは、この間の消息を端的に物語るものである。

スライディング制は勿論賃銀についてだけ考へられることではない。特に問題となるのは債權債務の關係についてゞあつて、インフレーション時代には貸主は價値の下つた貨幣で返却される

から、實質的に損をし、逆に借主はそれだけ儲けることになる。戰前一萬圓借りて立派な家を建てた人が、今なら洋服の一着も賣りとばせば一萬圓位ひ樂に返へせよう。不都合これより甚だしきはない。スライドの必要な所以であるが、その基準を物價指數に求むべきか、生計費指數または何等か他の標準に求むべきかは、一概には言明できない問題である。併し一般には物價指數が妥當とされてゐる。最も困難なのは國際間の貸借の場合で、政治問題にまで發展した例さへ少くない。

三　消費の緊急度

種々なる所得階級の家計支出を食費、衣服費、住居費等々に大別すると、勿論そのどの項目も所得の增加と共に增加することは勿論だが、英國の有名な統計學者ボーレイ及びアレンは、その何れもが、直線的增加であることを發見した。次圖は彼等が一九二七―二八年のドイツの資料から作製したものである。（Bowley & Allen, Family. Expenditure, 1935 p. 28）（第十六圖）

支出の順位は所得の大小によつて相違することが判る。例へば二千馬克階級では食料・其他費目、家賃、衣料、燃料、家具の順であるが、五千馬克階級では食料、其他費目、衣料、家賃、家具、

燃料の順となつてゐる。これらの事實から彼等は眞に必要な支出の順を定めんとし、上圖の如く、その直線（これは後に述べる回歸線である）を所得零の點まで延長した。所得が零なら勿論何も買へないわけで、順位もあつたものではないが、理論的には考へて差支へない。それによれば順位は食料、家賃、燃料、家具、衣料、其他費目となる。彼等はこれを緊急度順位（Order of urgency）と名づけた。最惡の事態に於て人が何を必要とするか、この順位はわれわれの常識とも合致するものである。

（消費緊急度）

（第十六圖）

右に於て支出増加は直線的だといつたが、それは或る程度の所得までに妥當することで、それ

以上の所得階級に於ては急に方向を變へるのである。エンゲル法則によつて食費の割合は遞減するから、食料の線は次第に傾斜が緩かになり、これに反して低額所得階級では殆どなかつた貯蓄が急激な勢で上昇する。アメリカの資料では、年收一六〇〇〇弗階級では貯蓄六〇〇〇弗、住居費三三〇〇弗、衣料費一四〇〇弗、食費一一〇〇弗で、低額所得階級とはまるでちがつた順位を示してゐる。

第十章　生産指數の諸問題

一　生産指數の擡頭

自由經濟の下に於ては、價格の優位に陰蔽されて、物量の重要さは不當に閑却されてゐた。經濟活動の彈條が利潤といふ貨幣的範疇に存する以上、人の注意が先づ價格現象に向けられるのは當然であるが、生産自體が自由に放任されてゐる場合には、生産は價格によつて規定されるから、換言すれば、生産は價格の函數として現はれるから、生産量は謂はゞ二次的地位を占めるに過ぎない。經濟のこの價格中心主義は經濟指數にも直接に反映し、從來の指數の殆ど全部が經濟の價格面の把捉に集中されたのである。卽ち經濟指數として直ちに何人の腦裡にも浮び來るものは、物價指數・生計費指數・賃銀指數・株價指數・生活用品小賣價格指數等々であつて、また經濟活動の消長を綜合的に表示せんとする景氣指數も亦、殆ど價格系列のみから構成されてゐるのである。

然るに例へば農業に於けるが如く、生産が主として自然的要因によつて規制され、一般の經濟

財と異つて自由生産の原則が著しく阻碍されるものについては、物量の重要さは最初から見失はれることはなかつた。これらの財に於ては、寧ろ價格は生産量の函數たるの性質を具へ、殊にそれが國民生活に必需の財たる場合には、物量に對する關心は一層高められざるを得なかつた。經濟統計の殆ど全部が價格統計たりし時代にも、主要農産物については生産量統計が明かに獲得されてゐたのであつて、本邦に於ても米麥收穫高は明治五年頃から發表されてゐる。

物量に對する關心は、しかし、自由經濟の凋落と共に急激に喚起されたのである。既に前大戰後の急激なる價格變動によつて、價格と物量との關係は著しく不明瞭となり、價格を通じて物量を推知することは次第に困難となつた。所謂貨幣的覆面をはいだところの物量そのものを直接に把捉せんとする希望は、この頃を境として頓に高まつたのであつて、生産指數が眞劍に考慮されるに至つたのもこの頃からである。然るに國際關係が再び攪亂され、各國とも平時態勢から準戰時態勢へ、更に準戰時態勢から戰時態勢へ移行するに及んで、價格經濟は根本から動搖して、軍需生産擴充を目標とする物資經濟が壓倒的優位を占めるに至つた。この國防經濟の下に於ては、價格と物量との關係は殆ど或ひは全く消滅し、一を以て他を推すことは不可能となつた。價格自體が全面的に否定されたのではないから、價格統計が存在理由を失つたといふのは明かに早計である。しか

しその自由作用が甚だしく失はれたことから、價格統計の意義が薄らいだことは確かであり、これと反比例して物量統計の意義が高められたことも亦確かである。斯くて從來不當に閑却された生産量指數が統計學の新たな課題として登場したのであつて、戰爭の生んだ一つの産物である。

生産指數の歴史は一九一三年に始まる。即ち同年レオナード(W. E. Leonard)が米國統計協會雜誌に發表した An Index of Changes in Extractive Industries を以て嚆矢とする。しかし實際に各國が之が作製に乘出したのは遙かに後のことで、殊に月次指數に至つては略々一九三〇年以降の産物である。我國の最初の試みは昭和二年(一九二七年)、當時名古屋高等商業學校講師たりしペンローズ(E. F. Penrose)の計算した農業生産指數(註一)であらう。同氏は二年後には鑛業生産指數(註二)を發表し、更に同氏の歸國後、同校調査室は昭和五年工業生産指數を(註三)、昭和八年農産物・畜産物・水産物・林産物・鑛産物及び製造品を一括する綜合生産指數を作製した(註四)。これには郡菊之助及び山田保治氏の大なる努力が秘められてゐると聞く。これら指數は何れも年次指數であつて、長期的動向を表示することを目的とする。月次指數としては三菱經濟研究所、ダイヤモンド社、東洋經濟新聞社及び商工省の工業及び鑛業生産指數があつた。

生産指數の最も發達してゐるのはアメリカで、主たるものは次の如くである。

Index of industrial production of the Federal Board.

Index of industrial production of the Standard Statistics Company.

Index of industrial production of Y. S. Leong.

Barron's index of crop production, industrial production, and trade.

Brookings institutions index of the composite physical volume of production.

Indexes of production and trade of the Federal Reserve Bank of New York.

（註一）本邦農産物の生産數量指數に就て、（名古屋高商産業調査室、調査報告第三輯）

（註二）日本鑛産物の生産數量指數（同第六輯）

（註三）本邦製造業の生産數量指數（同第九輯）

（註四）本邦生産數量指數總覽、自一八九四年至一九三一年（同第十四輯）

二　生産指數の種類

生産指數とは基準時點と比較時點との間に於ける生産量の比、即ち生産量の時間的變化率を示

す數字である。それは個々の生產物についても求められるし、またこれを一括した場合についても求められる。前者を個別生產指數、後者を綜合生產指數といふ。綜合生產指數は、それが同一產業に屬する生產物を一括した場合と、全產業を一括した場合とがあり、前者は所謂類別生產指數であつて、夫々の場合によつて農業生產指數、工業生產指數、鑛業生產指數等々と呼ばれ、後者は總生產指數と呼ばれる。更に比較すべき時點間の距離が一ケ月なるか一ケ年なるかで年次及び月次生產指數に分たれる。また算式に於ける加重の有無によつて加重生產指數と單純生產指數の別が生じ、平均の際に算術平均をとるか幾何平均をとるかによつて夫々の指數が區別される。

更にワーゲンフユールは別の見地から次の三つの型を區別してゐる（註二）。第一は、偶々存在する諸統計を平均して之に生產指數の名を冠したもので、彼はその適例として日本の三菱經濟研究所の舊指數を擧げてゐる。資料が極度に貧弱なる場合には、この方法によらざるを得ないが、その不完全なることは論を俟たない。第二の型は、主として景氣變動樣態を知る一手段として考へられたもので、景氣に最も左右される所謂敏感的な產業部門についてのみ求められた生產指數である。それが大なる利用價値を有することはアメリカに於て旣に立證されてゐるが、しかし國民經濟的意義に於ける變動率を正確に反映しうるか否かは素より疑問である。第三の型は、

彼の謂ふ「完全生産指數」vollständiger Produktionsindex であつて、全産業の生産變化率を、換言すれば、産業的實所得の變化を可及的忠實に反映せしめんとするものである。今日普通に言ふ意味の生産指數はこの型であつて、理論的にも技術的にも最も多くの困難を包藏してゐるのである。

　價格指數に於けると同樣、個別指數の作製は形式的には極めて簡單である。即ち特定生産物の基準時點の生産量を q_0、比較時點のそれを q_1 とすれば、$\frac{q_1}{q_0}$ が求むる指數である。故に信賴すべき統計さへあれば、この計算には何等の問題も介入し來らざる筈であるが、事實は q_0 及び q_1 の決定に多少の操作を必要とする。基準年度と基準年數の決定が、特に月次指數に於ては季節變動の取扱ひ方が問題となる。次に綜合指數に於ては原則として加重法がとられねばならず、從つて評量値の決定が重要問題となるのである。しかし一旦評量値 w', w'' …… w_n が決定されゝば、n箇の個別指數の綜合は、算術平均によれば

$$I_{01}=\frac{\frac{q'_1}{q'_0}w'+\frac{q''_1}{q''_0}w''+\cdots\cdots+\frac{q_1^n}{q_0^n}w^n}{w'+w''+\cdots\cdots+w^n}=\frac{1}{\Sigma w}\Sigma\left(\frac{q_1}{q_0}w\right)$$

幾何平均によれば

$$I = {}^{w'+w''+\cdots\cdots+w^n}\sqrt{\left(\frac{q'_1}{q'_0}\right)^{w'} \times \left(\frac{q''_1}{q''_0}\right)^{w''} \times \cdots\cdots \times \left(\frac{q^n_1}{q^n_0}\right)^{w^n}} = {}^{\Sigma w}\sqrt{\Pi\left(\frac{q_1}{q_0}\right)^{w}}$$

となる。

（註一）R. Wagenführ, Produktionsindexziffern und ihre Probleme (Vierteljahrshefte zur Wirtschaftsforschung, 13. Jahrgang, Heft 1)

三　資料の問題

有効なる統計的研究の不可缺前提が正確豊富なる統計資料に在ることは言ふまでもない。然るに從來の經濟統計が主として價格現象を對象とし、物量統計が著しく不備なることは既に冒頭に一言した。その際擧げた理由は、自由經濟の下に於ては、價格が經濟活動の規定者たることから、統計も亦自ら價格に集中されたといふことであつた。併しこの必要不必要といふ理由の外に、更に物量の統計化を阻碍し來つた別箇の事情がある。價格は單一共通の表現手段であるから、異種の生産物も、これを價格の觀點から眺めれば、全く同一の範疇に歸し、從つてこれを相互に加へ合はすことも自由である。斯かる統一性或ひは同種性こそ、事物の統計的把捉の必要條件であつて、これあるが故に、雜多な生産物も價格的にはこれを單一の統計數字を以て表示し得るので

ある。然るに物量は、これを表示すべき共通的尺度を缺き、從つて異種の生產物は多くの場合單一な數字的把捉を許さない。素より別箇の尺度を以て表示された物量も、簡單な換算によつて同一單位となすことが出來、延いて容易にそれらの總計を求めうる場合がある。噸と貫とで示された重さ、尺と米とで示された長さなぞこの例である。併し重さと長さを綜合することは不可能であつて、こゝに物量統計の最大の困難さがあるのである。

次に、同一單位を以て表明される同一生產物、例へば米穀をとつて見ても、事實はその內容必ずしも均一でなく、各種の品質、各種の銘柄から成つてゐる。工業製品に至つては、これは遙かに複雜である。この場合單に單位合計を以て果して滿足しうるであらうか。自動車月產二千臺とみ發表されたのでは、恐らく何等の知識も得られないであらう。この數字は乘用車・トラック其他各種自動車より成るが、その組合せの如何によつてこの數字の意義は著しく異つて來るであらう。假りにこれが更に分類されて、乘用車何臺、トラック何臺等と發表されたとしても、それらは各々多種多樣の內容を有するから、上記の困難は依然つき纏つてゐるのである。これを完全に克服せんがためには、眞に同一の生產物のみを集計する外はないが、これより生ずる種目の極端な細分は、それだけ多種なる統計系列を生み、人はその作製の煩に堪へざるのみか、たとへ作

製されたとしても、これが處理に當惑するに過ぎないであらう。

斯かる理由から、物量の數字的把捉は著しく制限され、それが絕對に必要なる場合にも、屢々より容易な他の方法によつて糊塗しつゝある狀態である。例へば昭和十四年の「物の國勢調査」は、國民の生活に直接必要なる物資につき消費の實狀を究めんとする目的に出でたものであるが、結局は物品販賣業者の賣上高又は接客業者の營業上必要なる物品の仕入高の調査に終り、僅少の例外を除いて、殆ど價格のみが調査された。要するに生產量統計が比較的容易に求められるのは、品種の比較的統一せられてゐる原始產業であつて、加工產業、就中規格の最も統一されざる手工業については、困難は特に大きいのである。

無數の生產物の一々については勿論、嚴密に言へばその大部分について、物量統計の存しないこと、乃至は利用し得ないことは上記の通りであるが、このことから、生產指數の不可能を論斷するのは早計である。指數の目的は變化の割合を近似的に把捉すれば足るから、第一には生產物の全種目を包含する必要はなく、第二には包含された種目の各々が完全無缺たるの必要もないのである。前者について見るに、一國生產活動に於て主要と認められるもの、即ち重要生產物のみを以て足ること、恰も物價指數が重要商品の價格のみから構成されて支障ないのと同じである。

商工省の製造工業生産指數は綿絲・銑鐵・機械器具等二十五種目、同じく鑛業生産指數は金・銀・銅・硫黄・石油・石炭の六種目のみから構成された。次に後者について見るに、或る計數が不完全なりとしても、その程度が略々一定ならば、比率を見出すためには差支へない筈である。工業製品の生産調査は一般に一定規模以上の工場（例えば職工數五人以上の工場）について行はれるから、總生産高を示し得ないことは事實であるが、包含される工場の生産高は總生産高の大部分を占めるのが通例で、從つてその代表力はいつも略々同じ程度に大きいと言つてよからう。併し包含される工場數が比較的に少い場合には、その生産高の總生産高に對して占める割合は著しく可動的となり、延いてこれに基いて作製された指數は信賴し得ざるものとなる。

如上の所論はたとへ不充分ながら資料の存在する場合についてゞある。然るに既述の如く、特に考慮に入れねばならぬ重要産物については動々もすれば資料を得難く、また或る種の例に於ては、生産量の資料は存在しても、これを指數の資料として利用することが不適當なのである。この最後に述べたものゝ例としては、完成に多大の時間を要する造船・建築・土木等を擧げることが出來る。斯かる事業の生産高は完成を俟つて始めて記錄される。もし生産指數の利用目的が生産物の變化率に在るならば、素より右の記錄によるべきである。併しもしその目的が生産活動の

變化率に置かれたとすれば、就中月次指數に於ては、斯かる記錄は目的に適はぬことになる。生産活動は久しく續行され乍ら、生産量としては遙か數ヶ月後に、且つ唯だ一回だけ、現はれるに過ぎないからである。

斯く資料の缺如し又は不適當なる場合には、これに代る資料に據らざるを得ない。代るといふのは、生産量又は生産活動と密接に關聯し、可及的に比例關係に立つものを指すのであつて、耕地面積・投下勞働量・機械運轉時間・原料又は燃料消費高・投下資本量・支拂賃銀額等は多かれ少かれこの性質を具へてゐる。しかし言ふまでもなく生産量は勞働・土地及び資本の複雜な函數であるから、それらの一つが生産量と正比例する如き場合は無いと言つてよいであらう。一般に勞働量は最もよくこの性質を具備するものと言はれるが(註一)、例へば我國の紡績業について見れば、昭和元年には男女工合計十八萬二千人强なりしものが、十年後の昭和十一年には十五萬二千人に減少し、而もこの間に生産量は二百六十萬梱から三百六十萬梱に、即ち一人當り一四・三梱から二三・七梱に增加してゐる。この場合には兩者の關係は全く逆であるから、勞働量を以て生産量の指標となすことは全く許されない。殊に昭和元年には男工は職工總數の二二・四%を占めたに對し、昭和十一年には僅かに一一・一%に減少してゐるから、質的にも明かに低下してゐる

のである。故にこれを計算に入れゝば、右の逆關係は一層甚だしくならねばならぬ。これに較べれば、運轉錘數は遙かによき指標と言へよう。即ち右十年間に五〇六四〇〇〇本から八四九八〇〇〇本に、即ち六八%の增加を示してゐるからである。しかし生產量の增加は三八%に過ぎないから、なほ甚大なる懸隔がある。原料消費量其他についても同じことが言へる。

併しこのことから代用系列の使用を全面的に否定するは當らない。紡績業に於ては昭和五、六年頃徹底的合理化が實施されたゝめ、勞働能率が急激に上昇し、以て上記の結果を來したのである。斯かる合理化は必ずしも總べての產業に行はれるものではない。農業に於けるが如く、たとへそれが行はれても極めて緩慢なる場合には、勞働量は依然生產量のよき指標である。周到なる研究が行はれるならば、恐らく如何なる產業部分についても相當信賴に値する代用系列は求められるであらう。

（註一）Bramstedt, Gefüge und Entwicklung der Volkswirtschaft (Allg. Statistisches Archiv. 25. Jahrgang. S. 398)

一般に指數の基準としては、或る時點の數量を常にその直前の時點の數量と比較するところの連鎖式と、一旦決定した基準をいつまでも踏襲するところの固定式とがあるが、物價指數に於けると同様、生産指數も實際には殆ど例外なく後者を用ひてゐる。前者は基準の變更が容易で且つ基準を常に更新するといふ利益があるが、同時に操作が煩雜たるのみならず、その具體的意味が不明瞭で、加之指數の誤差が累加されるといふ缺陷があるからである（註一）。併し固定式に於ても、單一年度を採る場合と（註二）、數ケ年を採るところの廣礎法とがある。本邦生産指數を見るに、三菱經濟研究所は昭和五年、ダイヤモンド社は大正十年を採り、東洋經濟新報社及び商工省は共に昭和六、七、八年を採つた。基準は元來正常値たることを要件とするから、數ケ年の平均値を採る廣礎法は、この點からより適當と認めねばならぬが、同時に一般に平均値に隨伴する不利、即ちそれが抽象的で不明確だといふ缺陷は避け難い。即ち基準生産量は數ケ年の平均であるから、基準年度の指數も一〇〇とはならず、一見して基準年度の所在を認識し難いことになる。

廣礎法をとる所以は數ケ年の平均によつて正常生産量を求めるに在るから、基準年數の決定が重要問題となる。これについては從來特別の考慮の拂はれることなく、單に數ケ年を平均すること

とによつてより正常的な値が得られるとの前提から、基準年度の前後數年（最短はブルガリアの二ケ年、最長はエストランドの五ケ年）を謂はゞ隨意に選定したに過ぎない。然るに山田勇氏は、その農業生産指數に於て、基準年數は經濟週期に等しからざる可らずとの理由から、シュスターの週期檢出法を適用して之が合理的解決を試みてゐる（註一）。これは正に指數論に於ける劃期的業蹟と言はねばならぬ。蓋し循環運動を有するものについては、週期年數に該當する數年間の生産量を平均したものは、最も妥當な正常値と考へられるからである。併しこの方法は、それ自體著しく週期的な自然的要因によつて規制される農産物については確かに効果的であるが、週期の曖昧な工業又は鑛業の諸産物については必ずしも適用し得ないであらう。

注意すべきことは、物價指數に於ては基準の正常性が特に要求されるに對し、生産指數に於ては寧ろ最も近い過去に於ける好況期を採るべしといふ意見のあることである。これは、生産は物價と異りその大なるを理想とするからで、右の方法によれば生産恢復の情況が明示されるであらう。各國の指數に一九二八年の基準が最も多く見受けられるのはこの理由によるものと考へられる。併しこの理を推し進めれば、計畫經濟の下に於ては、或將來年度を基準とすることも可能であらう。例へば産業五ケ年計畫が實施される場合、生産豫定量の實現するであらう五年後を基

準とするのであつて、豫定量への年々の接近程度は之によつて明瞭となろう。

基準の決定に伴ふ諸困難を克服せんがため、紐育聯邦準備銀行は一九三八年新たに作製した Indexes of production and Trade に趨勢値基準とも稱すべき新方法を採用した（註四）。即ち指數の內容をなす各系列につき趨勢値（註五）を算出しこの推定趨勢値（Estimated longterm trend）を以て基準となすのである。趨勢値は正常値の軌跡或ひは動的正常値と認めれるから、この方法は固定基準と連鎖基準の兩者の特長を兼ねたものと見られ、理論的には極めて巧妙な構成と言はざるを得ない。但し趨勢値なる概念が全く專門的なることを一考すれば、この種の指數の利用範圍の著しく制限されるであらうことは免れない。

最後に基準に關聯して季節變動の問題がある。年次指數は、年生產高に基いて計算されるから、季節の要素は介入しないが、月次指數に於てはこの要素は時に極めて重大なる問題となるのである。日次指數は或る月の生產量を基準時點の同じ月の生產量と比較することによつて求められねばならぬ。蓋し自然的條件及び日數（詳しくは勞働日數）の不一致によつて、每月の生產量は當然相互に異るからである。故に基準數量として、一ケ年又は數ケ年の單純なる月平均を採つたならば、比較時點の月產量は、これに對應しうるが如き修正を必要とする。もしこの修正を免

かれんとすれば、即ち與へられた月産統計をそのまゝ使用せんとすれば、豫め基準數量がこの季節的修正を受けて居らねばならぬ。勞働日數の多少より生ずる變化は、月産量の代りに一勞働日當り生産量を使用することによつて、自然的條件に基く變化は、季節指數の算出によつて、はじめて修正が可能となる。商工省の工業生産量指數は、基準として月當り平均生産高をとり、比較數量に關しては、「季節的變動及び作業日數の多少に基く影響に對しては修正を施さず」と註されてゐるから、利用に當つては戒心を要する。東洋經濟新報社指數は季節變動調節指數である。

（註一）山田勇、東亞農業生産指數の研究、二七六頁。

（註二）那國の舊指數は半ケ年（一九三三年前半期）を基準とした。

（註三）山田勇、前掲書第五章及び數學附錄

（註四）N. O. Johnson, New Indexes of production and Trade (Journal of the American Statistical Association, vol. 33. No. 202.)

（註五）主として用ひられた趨勢線は $Y=bc^{1/(a+x)}$

五　綜合の方法

綜合指數の作製方法としては、總和法と平均法との二種がある。いま物價指數について見れば、總和法は、選定された品目の基準價格の總和を以て、比較價格の總和を除した商を求めることである。平均法は、個別價格指數を平均する方法で、その場合、算術平均をとるか幾何平均をとるかによつて二つに區別される。

しかしこれら諸式は、各品目の重要度は何れも等しいといふ假定に立つもので、事實に於ては成立し得ない。品目による重要度の相對的大さを示す指標をウエート又は評量値といひ、これを算入した算式を加重算式といふ。上記の物價指數に於ては、總和法では各商品の取引數量が、平均法では取引價格が、それぞれの品目の評量値となることは既述の通りである。評量値をWとすれば、上記三式は夫々

$$P_{01} = \frac{\Sigma(p_1 w)}{\Sigma(p_0 w)}$$

$$P_{01} = \frac{1}{\Sigma w}\,\Sigma\left(\frac{p_1}{p_0}\cdot w\right)$$

$$P_{01} = \sqrt[\Sigma w]{\Pi\left(\frac{p_1}{p_0}\right)^{w}}$$

となる。そして總和法のwはq、平均法のwはpqであるが、p及びqは基準時點と比較時點でそれぞれ異るから、總和法のwはq_0又はq_1、平均法のwはp_0q_0、p_0q_1、p_1q_0又はp_1q_1となる。但しその何れか一つを使用すれば、所謂偏倚 bias の介入し來る懼れがあるため、相對立する性質の評量値を平均する方法、例へば

$$P_{01}=\frac{\Sigma\frac{q_0+q_1}{2}p_1}{\Sigma\frac{q_0+q_1}{2}p_0}=\frac{\Sigma(q_0+q_1)p_1}{\Sigma(q_0+q_1)p_0}$$

又は相對立する評量値による二つの算式を平均する方法、例へば

$$P_{01}=\sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0}\times\frac{\Sigma p_1q_1}{\Sigma p_0q_1}}$$

の如きものが考へられる。前者はエッヂワース式又はボーレイ式、後者はフィッシャー理想式と稱せられる。

加重を度外視すれば、綜合生産指數は單純算式による物價指數式の價格の代りに數量を、即ちpの代りにqを置けばよい。但し數量の表示に異る尺度の用ひられる品目については、單純總和法の用ひうべからざることは勿論である。しかしこの障碍も、適當なる評量値を與へることによ

つて克服されるし、事實加重法によらざる生産指數は先づ考へられないから、問題は生産指數に於ける評量値は如何にして決定されるかといふことになる。

物價指數に於ては評量値は、上記の如く、總和法では取引數量、平均法では取引價格である。この簡單さは物價指數が貨幣價値變動の測定といふ明白な目的を有するに出づるのである。生産指數は一般的に言へば生産量變動を測定するものであるが、これは畢竟實物經濟の活動程度を知らんがためである。然るにこのことはそれ自體著しく曖昧な概念であるから、多様の解釋が可能とならう。別の言葉で言へば、評量値決定の基礎は、物價指數の場合と異つて、幾通りもありうるといふことである。いま各國の生産指數を見るに、評量値として勞働者數、投下資本額、利潤、使用動力、生産價額の如きものが用ひられてゐる。このうち生産價額は總生産額（Bruttoproductionswert）と純生産額（Nettoproductionswert）とに分たれる。前者は原料費その他の生産費を含めたもの、後者はこれらを控除した所謂「添加價値」（value added）である。

これら多種多様な評量値は、素より何れも各生産部門の重要性を反映する指標として選ばれたわけであるが、それらの經濟的意義の相互に甚だしく相違する事實を顧れば、生産指數に於ける評量値の決定が如何に困難なるかは明かであらう。勞働者數や投下資本額等が生産數量と比例す

るか否かは、既に代用系列の問題として説明したが、こゝではこれらが異る生産部門の相對的重要性と比例するか否かが檢討されねばならぬ。いま勞働者數を評量値とした場合を考へれば、これは凡ゆる生産部門を通じて勞働者一人當り能率が等しいといふ假定の下に於てのみ妥當する。特定部門のこの能率が時間的に變化するといふ事實あるがために、勞働者數を以て生産量統計に代へることが困難だといふ次第は既に述べたが、評量値としてならば、その時間的變化そのものは格別支障とはならぬ。要は如何なる時點をとつても、そのときの勞働能率が何れの部門に於ても等しければよいのである。しかし改めて説明するまでもなく、斯かることはあり得ない。要求される勞働熟練度は産業によつて極度に相違するからである。農業に於ては、この點は比較的均一であるが、他方その勞働は一生産物に集中されないため、箇々の農産物が幾何の勞働を必要としたかは計算し難いといふ缺陷がある。即ち勞働者數は一般に評量値として不適當であり、強いて用ふれば、勞働生産力の低い生産部門ほど相對的に大なる重みを與へられることになるのである。投下資本量、使用動力量その他何れも、極めて有り得ざる假定の下に非んば、安んじて評量値と爲すことは出來ない。

斯くて結局最も無難なものとして、廣く生産價額が用ひられるのである。價格經濟の下に於て

は、各産業部門の重要性はその與へる生產價額によつて最もよく測られるであらう。こゝで問題となるのは、價額として總生產價値をとるか、純生產價値卽ち添加價値をとるかであるが、これは產業の性質によつて決定される外はないであらう。商工省指數では、「評量値の決定に際しては原則として製造工業に付ては當該生產過程に於ける添加價値を基準とし、鑛業に付ては總生產額を基準とす」と記されてゐるが、一般に加工產業たる製造業にあつては、製品の價値は大なる程度に原料價値によつて占められるから、總生產額をとれば明かに重複計算となり、當該產業の重要性は過當に評價されることにならう。鑛業又は農業の如き原始產業に於ても、このことは或る程度までは當嵌まるが、その占める割合の一般に低いこと及び明瞭ならざることから、概して總生產額が用ひられるのである。

六　品質の問題

生產指數に隨伴する困難は、上記の如く多々あるが、最近特に顯著となつた別箇の問題がある。生產物の品質及び利用法に於ける變化がこれである。同一名稱を以て呼ばれながら、事實は多種多樣な品種銘柄より成る生產物が、少くないといふよりは寧ろ原則であつて、このことが生

產量の數字的把捉を困難ならしめてゐることは前述したが、斯かる相異は同一物について時間的にも起りうる。從來主として問題とされたのは、時の經過による品質及び利用技術の改善であつて、これを度外視した生產指數は、比較時點の實狀を過當に低く示す傾きがあるといふことであつた。米國のポートランド・セメントの一九一八年の生產量は七一〇〇萬俵であつたが、一九三六年には一一二〇〇萬俵へ、即ち五八%、或ひは年平均二・六%增加した。然るにこの間に品質と利用技術は飛躍的に改善され、同量のセメントは二倍の効果を生むに至つた。故にこの事情を考慮に入れゝば、一九三六年の生產量は二二四〇〇萬俵に該當し、增加率は二一五%、即ち年平均六・六%と改められねばならぬ（註一）。進步的經濟の下に於ては、このことは大部分の製品及び原料について言ひうることである。唯だ農產物については、品種改良は勿論絕えず行はれてはゐるが、他に較べてその程度は少いと思はれる。農業生產指數が比較的容易に求められる理由の一つはこゝに在る。

然るに鑛產物については、屢々時の經過と共に逆に品質の低下が認められる。これは一方では製鍊法の改善によつて、從來顧られなかつた貧鑛に對する需要が增大した爲であるが、これを別としても富鑛が次第に採掘され盡すといふ不可避的事情があるからである。併しながら、原則と

しては次第に品質の向上すべき工業製品についても、最近では品質及び耐久力の低下が顯著たらんとしつゝある。資材、勞力の缺乏及び價格統制の當然の結果である。或る製品の耐久力が半減したとすれば、上記のセメントの場合と反對に、生產量は實質的には半分として計算さるべきであらう。

しかし品質の變化は、改善であれ低下であれ、何れも正確な測定は不可能であらう。然らばこれを考慮して生產量に修正を加へることも亦不可能であつて、事實斯くの如き修正を經た生產指數は未だないと思はれる。これと關聯して一考せらるべきは、所謂代用品の取扱ひ方である。從來の種目と同一視すべきか、或ひは全くの新種と見るべきか。恐らくこれに對して一般的解答を與へることは出來ないであらう。それは原料、用途範圍及び耐久力等の相異程度によつて、或るものは舊種へ接續されてよく、或るものは新種と見做されねばなるまい。もし後者と認められたならば、追加種目として生產指數にとり入れるべきか否かゞ改めて檢討されねばならぬ。そしてもしとり入れたとすれば、指數本來の特色たる前後比較性がこれによつて妨げられぬよう、可及的注意を必要とする。

（註1）P. G. Hudson, The Technical problems and limitations to the construction of indexes

of Physical Production (Journal of the American Statistical Association, Vol. 34, No. 206, pp 249-250.)

七　結　　論

生産指數が生産量の眞の變化率を表示せんとしながら、實際にはこれを或ひは不當に高く、或ひは不當に低く表示せざるを得ない原因として、ハドソンは次の如き一覽表を揭げてゐる。これは本稿で取扱つた諸理由を略々網羅してゐると思はれる。

一　上向的偏倚を惹起す諸要素

（1）若干の採取產業よりの產物及び、或る場合には、工業製品の品質の漸次的低下

（2）使用された算式の型と選定された基年より來る數學的の上向的偏倚

（3）資料蒐集の完備さの增大

二　下向的偏倚を惹起す諸要素

（1）農產物の品質の漸次的改良と、多數工業製品の品質の著しき改良

（2）同量の原料よりより多くの使用價値の抽出を可能ならしめる進步的利用法

（３）副産物及び急激に發展しつゝある新興産業を指數作製にとり入れざりし場合

（４）使用せる算式の型と選定された基年より來る數學的の下向的偏倚

これら諸要素の多くは一つの生産指數のなかに同時に介在しうるから、問題は、これら相反する性質の何れが全體的に見てより優勢かといふことである。素よりこれは一々の實例について答へられるべきことで、先驗的に斷定すべき性質のものではないが、ハドソンは下向的偏倚のより優勢なることを暗示してゐる。即ち彼に從へば、發表された生産指數が正(プラス)の增加率を示すときは、生産實情はより一層大なりと見てよいと言ふのである。しかし今日の如く、品質の犧牲に於て、或ひはより劣等な代用品の奬勵によつて、量の增加を計る實例の乏しからざるを一考すれば、右とは逆の結論も可能であらう。

何れにしても、眞に信頼しうる生産指數が望み得べからざるものとすれば、過度に綜合的な生産指數は寧ろ避くべきである。箇々の類別指數すら極めて不完全なる場合、これらを網羅せる全生産指數を求めんとするが如きは、危險極まる企てである。先づ要請せらるゝものは、可及的正確なる類別指數、即ち農業・鑛業・工業別生産指數である。これら産業は各々獨自の組織をもつから、資料蒐集の方法、基準の年度及び年數、評量値の決定等に於て一律の扱ひ方は許されない

筈である。そしてこの理を更に押進めれば、これらの類別指數自體が更に小類別されねばならぬ。蓋し斯かる小類別によつて産業の具體的把捉が可能となるのみか、これらを類別指數に綜合するに際して更に改めて適當な評量値（小類別評量値）を與へることにより、更には類別指數をより廣範圍なる綜合指數に綜合するに際して更にもう一度適當な評量値（類別評量値）を與へることによつて、——或ひはその都度算式までも變更することによつて、——單純な綜合より生ずる不合理を大なる程度に排除しうるであらう。ワーゲンフェールは工業生産指數の小類別として、（一）食料及び嗜好品（註一）、（二）彈性的需要の消費財（註二）、（三）資本財（投資財）（註三）（四）其の他の生産財（註四）に分類した。

（註一）麵麭、小麥粉、肉、麥酒、煙草等。

（註二）織物、家具什器、玩具、樂器等。

（註三）機械、建物、原料等

（註四）石炭、瓦斯、電氣、用水、化學製品等。

その理由として彼の擧ぐるところは、この四項目は次の諸點に於て獨自の特徴を有するといふのである。即ち、生産量變動率より見れば（一）は弱、（二）、（三）は强、（四）は中庸、最

重要な販賣市場より見れば（一）と（二）は消費者（三）は生産者（四）は全經濟、販賣を決定する最重要要素より見れば（一）は消費者數、（二）は消費者所得、（三）は投資需要（資本投下量）、（四）は全經濟、販賣を決定する最重要要素より見れば（一）は消費者數（二）は消費者所得、（三）は投資需要（資本投下量）（四）は一般經濟活動である。なほ先きに言及した紐育聯邦準備銀行の Indexes of production and Trade は、その表題の示す通り、純粹な生產指數ではなく Production, Primary Distribution, Distribution to Consumer, Miscellaneous services の四大項目より成るが、全評量一〇〇のうち「生產」の占める評量値は五五に達し、系列數に至つては、全數八二中六一を占めてゐる。卽ち生產に重點を置いたことは明かであるが、いまその部門の構造を見るに、Producers' durable, Producers' nondurable, Consumers' durable, Consumers' nondurable goods, Employee-hours の五部に分たれ居り、從つて類別生產指數として（一）耐久生產財、（二）非耐久生產財、（三）耐久消費財、（四）非耐久消費財の四指數が得られることになる。（一）（二）を合はせて系列數三〇、評量値二三・二、（三）、（四）の合計も亦同じである。

戰時には戰爭目的のため重視された生產指數は、今日では復興目的のための缺く可らざる指標

となつた。生産の恢復以外にインフレーションを克服し經濟危機を打開する途のないことは言ふ迄もないことで、今日ではズブの素人さへ石炭の、電力の、肥料の生産狀況に一喜一憂の有様である。併し綜合生産指數は未だ不完全極まる。經濟白書の資料となつた國民經濟研究協會の工鑛業生産指數は、昭和十年から十二年迄の平均を一〇〇とする加重算術平均式で、評量値は昭和十二年の生産價額である。試みにこれを列記すれば、製造工業（一〇〇）、纎維工業（三二）、化學工業（二一）、鐵鋼業（一四）、機械工業（一八）、窯業（三）、製造食品工業（一二）、鑛業（一〇〇）、非鐵金屬精錬（三八）、石油（三）、石炭（五九）である。綜合指數は終戰當時の八・七を最低とし、幾分上昇したとはいへ、なほ三〇臺を彷徨しつゝある狀態である。現在用ひられてゐるその他の工鑛業生産指數は安本統計課、ダイヤモンド社、東洋經濟新報社（月別—舊、年別—新）、等である。G・H・Qでも別に作つてゐる。

第十一章　發展の一般的方向（傾向線）

一　時系列解析の課題

萬物は流轉するとか、榮枯盛衰は世の慣ひといつた諺のとほり、總べては不斷に變化の過程を辿つてゐる。變化には極めて緩慢なものと急激なものとがある。地球の溫度は冷却してゐるといふが、それは何千年何萬年について始めていへることである。ところが紙を燃せば、見る見るうちに灰になつて了ふ。經濟現象についても同じこと。封建主義が資本主義に變るには可成りの年月を要したが、株式市場の相場は一日のうちに何度でも變つて了ふ。そして同じ現象でも時代によつて變化の緩急に甚だしい相違がある。物價は長い間比較的に安定してゐることもあり、最近のやうに僅かの間に數倍にはね上ることもある。時の經過に伴ふかゝる變化を數字的に把捉することは、同一時點に於ける事物の相違を數字的に把捉することゝ相俟つて、統計學に課せられた當然の任務である。

時間的變化を、一般に動態といふ。それが如何なる形で現はれるかといふに、次の三つの何れ

かである。第一は長期傾向または趨勢と稱する、長い期間に亘つて一定の方向に動く運動である。長いといふのは勿論相對的な意味で、決して嚴格な標準があるわけではない。「地球は次第に冷却する」といふ場合には、非常な長期を言つて居り、「近頃死亡率は減つてゐる」といふ場合には、僅か五年か十年位を指してゐるのである。第二は循環的運動で、或る期間がたてば元へ歸つてくるものをいふ。元へ歸るまでの期間は周期である。そして周期が確定してゐるかどうかに從つて、季節變動と景氣變動とに分ける。第三は右の何れにも該當しない、謂はゞ突發的變化で、これを不規則變動といふ。地震や戰爭等の突發的事故またはかゝる事故から起る諸結果、例へば地震による物價騰貴や戰爭による勞働力缺乏などこれに屬する。

統計集團の時間的變化を時間の順序に從つて配列したものが時系列である。故に時系列は上記の運動を示すものであるが、併し現實に與へられた時系例は、これら諸運動の何れか一つをも明瞭に示してはゐない。何となれば、いかなる事柄もこれら諸運動の全部乃至大部分を同時に行つてゐるからで、一つの時系列に現はれた變化は、畢竟これら諸運動の合成物なのである。純粹な季節變動だけを示してゐる時系列とか、純粹な長期傾向を示してゐる時系列とかは、先づ全然ないと言つてよい。然るに運動にかゝるいくつかの形があるのは、その各々の發生理由がちがふか

らである。理由がちがへばそれを區別する必要があるのであつて、それが科學的たる所以に外ならない。與へられた一つの時系列に分析的操作を加へて、以てそこに現はれた合成的變化をいくつかの基本形に還元することは、斯くて統計學の一つの課題である。これを時系列の解析といふ。

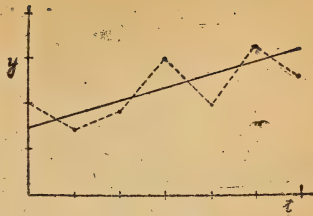
先づ長期傾向の問題から始めよう。

二　最小自乘法による傾向線の決定

(1) 直　線

長期傾向は一般に直線または拋物線のような簡單な形で示されねばならぬ。何となれば長期傾向とは長期に亘る變化の一般的方向を意味するからで、もし複雜な曲線で示せば、一般的方向は判らなくなつて了ふ。千鳥足の醉漢さへ、歩行の一般的方向は矢張り直線といへよう。そこで、長期傾向を求めるとは、與へられた時系列に直線または拋物線のような簡單な線を當嵌めることである。先づ直線から。

一つの時系列をグラフにして第十七圖の點線を得た。tは年次を示す。その一般的方向は、實

線の示す直線である。この實線はどうして得られるか。

最も簡單には目分量で（即ち目ノコ又は目測で）描けばよい。人の目は案外たしかなもので、少し慣れゝば、次に述べる數學的方法によつたものと殆どちがはぬ結果を得られるものである。そして既に述べたように、直線は必ず $y=a+bx$ といふ x に關する一次式で示されるが、a は直線と y 軸との交點の座標、b は x の一單位の増加に伴ふ y の増加（即ち直線の高まり方）であるから、圖の點線について、目盛に從つて數學的に求めることができる。

第十七圖

併し目分量は、慣れゝば上手になるとはいへ、元々主觀的判斷によるから、非常な正確さは期待できない。合理的で客觀的な方法が要請される所以であるが、これに應ずるものが、最小自乘法に基く數學線の當嵌である。上圖のように平面上にA、B、C……Nの n 箇の點が散在してゐるとき、これら諸點に最も近い直線は、これら諸點からの距離 d_1 d_2……d_n の自乘（平方）の總和 $d_1^2+d_2^2+\cdots\cdots+d_n^2$ を最小ならしめる直線である。Aの座標を（x_1，y_1）、Bのそれを（x_2，y_2）等々とすれば、

$d_1 = y_1 - (a + bx_1)$

$d_2 = y_2 - (a + bx_2)$

……………

……………

である。故に自乗の總和をSとすれば

$$S = [y_1 - (a + bx_1)]^2 + [y_2 - (a + bx_2)]^2 + \cdots\cdots + [y_n - (a + bx_n)]^2$$

となる。これを最小ならしめるには、微分法の原理に從つて、微係數 $\frac{\partial S}{\partial a}$ 及び $\frac{\partial S}{\partial b}$ を零ならしめればよい。然るに

$$\begin{cases} \frac{1}{2} \cdot \frac{\partial S}{\partial a} = \Sigma[y - (a + bx)] = \Sigma y - na - \Sigma x \\ \frac{1}{2} \cdot \frac{\partial S}{\partial b} = \Sigma[y - (a + bx)]x = \Sigma xy - a\Sigma x - b\Sigma x^2 \end{cases}$$

であるから($\Sigma a = na$ である)、これから次の聯立方程式が得られる。

$$\begin{cases} \Sigma y = na + \Sigma x \\ \Sigma xy = a\Sigma x + b\Sigma x^2 \end{cases} \quad \cdots\cdots(1)$$

これからa及びbを求めて、その値を直線方程式$y=a+bx$に代入すればよいのである。時系列では時の經過をx軸に示すのが習慣であるが、このことから計算上次の二つの便宜が得られる。抑も時系列とは連年または連月の統計であるから、時間の間隔は一ケ年または一ケ月で、その長さは等しい。（詳しく言へば必ずしも等しくはない。一年は必ずしも三六五日でなく、一ケ月必ずしも三〇日ではない。併しその差はこゝでは無視するのである）。そこで與へられた時系列の時を示す數字は、これを簡單な形に書き換へることができる。1930, 1931, 1932……これを0, 1, 2……としても差支へないのであつて、計算は甚だ樂になる。併しこの方法を更に徹底させ、もし與へられた年數が奇數のときは、中央の年を基準即ち零として書換へれば計算は一層樂になる。何となれば1930, 1931, 1932, 1933, 1934は−2, −1, 0, +1, +2となり、その合計は零となる。即ち上記の聯立方程式のΣxは零となり、次の簡單な形に變つて了ふからである。

$$\begin{cases} \Sigma y = na \\ \Sigma xy = b\Sigma x^2 \end{cases} \quad \cdots\cdots(2)$$

年數が偶數なら、中央はないから、−1.5, −0.5, +0.5, +1.5とするか、または−3, −1, +1, +3とする。後の場合では一年は2といふ長さで示されるわけである。何れにしろxの合

計が零となるようにすればよいのである。同一資料について、普通の方法と簡便法とによる計算實例をあげておかう。

この方程式のXに零を代入すれば一九二一年の傾向値が求められ1を代入すれば一九二二年のそれが求められる。これを全部について求めたものが次表に示されてゐる。

	X	Y (單位千円)	X Y	X^2
1921	0	16.82	0	0
1922	1	20.73	20.73	1
1923	2	26.68	53.36	4
1924	3	28.43	85.29	9
1925	4	31.94	127.76	16
1926	5	37.76	188.80	25
1927	6	41.38	248.28	36
1928	7	42.54	297.78	49
1929	8	48.45	387.60	64
1930	9	54.09	486.81	81
1931	10	41.67	416.70	100
合　計	55	390.49	2,313.11	385

$$\begin{cases} 390.49 = 11a + 55b \\ 2,313.11 = 55a + 385b \end{cases}$$

これを解いて

$$\begin{cases} a = 19.104 \\ b = 3.279 \end{cases}$$

∴求むる方程式は

$$Y = 19.104 + 3.279X$$

年次	傾向値
1921	19.104
1922	22.383
1923	25.662
1924	28.941
1925	32.220
1926	35.499
1927	38.778
1928	42.057
1929	45.336
1930	84.615
1931	51.894

X	x	Y	x Y	x^2
1921	-5	16.82	-84.10	25
1922	-4	20.73	-82.92	16
1923	-3	26.68	-80.04	9
1924	-2	28.43	-56.86	4
1925	-1	31.94	-31.94	1
1926	0	37.76	0	0
1927	1	41.38	41.38	1
1928	2	42.54	85.08	4
1929	3	48.45	145.35	9
1930	4	54.09	216.36	16
1931	5	41.67	208.35	25
	0	360.49	360.66	110

$$\begin{cases} 390.49 = 11a \\ 360.66 = 110b \end{cases}$$

これを解いて

$$\begin{cases} a = 35.499 \\ b = 3.279 \end{cases}$$

∴求むる方程式は　$Y = 35.499 + 3.279x$

この二つの結果は全く同じである。但し一方は一九二一年が、他方は一九二六年が基準であるから、y=a+bx のaの値は當然變つてくる。そのため一方では a=19.104 他方では a=35.499 となつたのである。

(2) 拋物線

曲線の最も簡單な形は拋物線で、その方程式は

$$y=a+bx+cx^2$$

なるxに關する二次式である。これを求めるにも原理的には直線の場合と同一で、次の聯立方程式から媒介變數a、b、cを決定し、それらを右の方程式に代入すればよい。

$$\begin{cases} \Sigma(Y)=Na+b\Sigma(X)+C\Sigma(X^2) \\ \Sigma(XY)=a\Sigma(X)+b\Sigma(X^2)+C\Sigma(X^3) \\ \Sigma(X^2Y)=a\Sigma(X^2)+b\Sigma(X^3)+C\Sigma(X^4) \end{cases}$$

但しこの場合にも、年次の中心を原點とすれば、$\Sigma(X)=0$, $\Sigma(X^3)=0$ となるから、次の簡單な形となる。（第十九圖）

$$\begin{cases} \Sigma(Y)=Na+C\Sigma(x^2) \\ \Sigma(xY)=b\Sigma(x^2) \\ \Sigma(x^2Y)=a\Sigma(x^2)+\Sigma C(x^4) \end{cases}$$

次にその應用例を示さう。

拋物線的傾向線

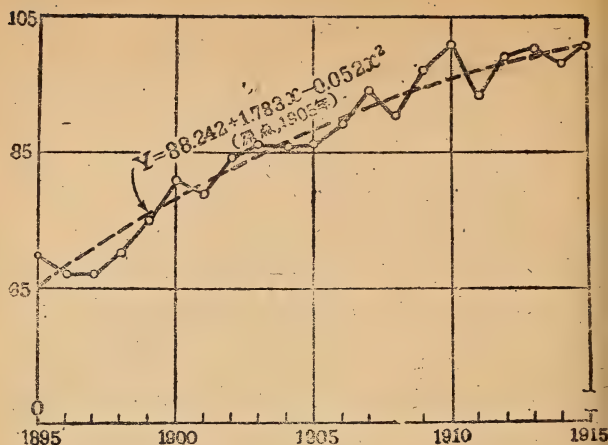
X	物價指數 Y	x	xY	x^2	x^2Y	x^4
1895	70	−10	−700	100	7,000	10,000
1896	67	−9	−603	81	5,427	6,561
1897	67	−8	−536	64	4,288	4,096
1898	70	−7	−490	49	3,430	2,401
1899	75	−6	−450	36	2,700	1,296
1900	81	−5	−405	25	2,025	625
1901	79	−4	−316	16	1,264	256
1902	84	−3	−252	9	756	81
1903	86	−2	−172	4	344	16
1904	86	−1	−86	1	86	1
1905	86	0	0	0	0	0
1906	89	1	89	1	89	1
1907	94	2	188	4	376	16
1908	90	3	270	9	810	81
1909	97	4	388	16	1,552	256
1910	101	5	505	25	2,525	625
1911	93	6	558	36	3,348	1,296
1912	99	7	693	49	4,851	2,401
1913	100	8	800	64	6,400	4,096
1914	98	9	882	81	7,938	6,561
1915	101	10	1010	100	10,100	10,000
	1813		1373	770	65,309	50,666

$$b=\frac{\Sigma xY}{\Sigma x^2}=\frac{1373}{770}=1.783$$

$$c=\frac{n\Sigma x^2Y-\Sigma x^2\Sigma Y}{n\Sigma x^4-\Sigma x^2\Sigma x^2}=\frac{1,371,489-1,396,010}{1,063,986-592,900}=\frac{-24,521}{471,086}=-0.052$$

$$a=\frac{\Sigma Y-c\Sigma x^2}{n}=\frac{1813-(-40.0801)}{21}=88.242$$

∴求むる方程式は　$Y=88.242+1.783x-0.052x^2$　（原點，1905年）

三　複雜な傾向線

傾向線は發展の一般的方向を示すのが目的であるから、複雜な高次曲線は不適當である。即ち直線か抛物線で大部分の場合充分間に合ふ。勿論特殊の目的に對しては可成り複雜な曲線も必要とならうが、それらは入門書では取扱へない。特殊な曲線とは例へばゴムペルツ曲線 (Gompertz curve) やロヂステイク曲線 (Logistic curve) の如きものである。この二つは共に事象の發展法則を示さんとするもので、主として人口增加の形態として一部の人口

第十九圖

學者によつて主張されたものである。ゴムペルツ曲線は $y=ab^{cx}$ 即ち

$$\log y=\log a+cx\log b$$

なる幾分複雑な形をとるが、その意味は、增加力は次第に衰へて最後には飽和點に達するといふことである。ロヂスティク曲線はこれを更に複雑化し、飽和點に達するまで種々なる形の增加力が現はれることを示すもので、その方程式は次の如くである。

$$Y=\frac{L}{1+e^{\frac{\beta-t}{a}}}$$

この曲線はケトレーがマルサスの人口原理に數學的表現を與へんとした努力に刺戟されて、フェルフュルスト（Verhurst）が始めて案出したものであるが、その後殆ど學界から忘れられてゐたものを、一九二〇年に至つて米國のパール（Pearl）及びリード（Reed）兩氏が再發見し、所謂ロヂスティク人口理論として特殊の地位を占めてゐるのである。

第十二章　季節指數

一　季節と季節指數

季節ほど人間生活に大きな影響を與へるものはない。スポーツについて見ても、野球、水泳、スキー、何れも獨自のシーズンをもつてゐる。經濟生活に於ては季節の影響は正に壓倒的といへる。夏ともなれば浴衣や氷が欲しくなり、冬に向へば外套や燃料が心配になる。併し季節の特に基本的な作用は農業といふ基本産業に最もはつきり現はれてゐるのである。殆ど凡ゆる農産物は一定の播種期と收穫期をもつてゐる。農村ほど暦を必要とするところはない。そして農産物價格は出廻期には下り、端境期には騰貴する。

季節は一年を週期として規則正しく循環してゐるから、季節に支配される人間生活にも一定のリズムが生れざるを得ない。勿論文明の進歩と共に人が次第に季節の束縛を脱してきたことは事實である。暖房、冷房の裝置によつて人は温度の變化から免れ、人工栽培によつて眞冬に苺にありつくこともできる。併しかような裝置を必要とすることが、取りも直さず人がいつの世にも季

節の支配下にある何よりの證據である。

經濟政策の見地からは、季節の及ぼす影響を正確に知ることが特に肝要である。失業は冬期に最も起り易い。季節の關係から農業、建築業、交通業等々が萎縮するからであるが、これが毎年のこと〻あらば、これを充分考慮に入れて失業對策を樹てる必要がある。また櫻の咲く頃ともなれば交通機關は必ず混雜する。交通會社はこれを勘定に入れて豫め準備するところがなければならない。最近は電力不足が甚だしいが、渇水期は毎年ほゞ決つてゐるのであるから、出來れば火力發電の用意をするとか、木炭の配給を増すとか、豫め對策があつて然るべきである。

併しそのためには、季節の影響を數字的に算出することが先決要件である。統計學の一つの課題は、この要求に應じうる季節指數を作製することである。それは既述の傾向値の決定と相俟つて、時系列解析の主たる内容をなすものである。

本來の季節即ち一年を週期とする規則的運動は月別統計の上に現はれる。季節の變化は月から月への推移に外ならぬから。いま月別統計から季節指數を求める方法を考へるに、對象が自然的現象であるか社會的現象であるかによつて甚だ性質の異ることが判る。前者即ち濕度や雨量などは、勿論年によつて相違があり、暑かるべき夏に濕度が低かつたり、雨の多かるべき梅雨期が空

つゆに終ることも珍らしくないが、元々規則正しい自然の運行に源を發する現象であるから、相當の長期間について平均的に觀察すれば、かような異變は相殺されるであらう。涼しい八月もあるが、何十年振りの暑い八月もあるわけで、平均すれば八月の正常溫度が判る筈である。よつていま相當長期間の月別統計を月の順序に配列し、一月の平均値、二月の平均値……十二月の平均値を求める。これを x_1, x_2……x_{12} とすれば、これは各月の正常値と認められる。これを指數に換算すればよいわけで、それにはこれら十二の平均値$\bar{x}$を求める。即ち

$$\frac{1}{12}(x_1+x_2+\cdots\cdots+x_{12})=\bar{x}$$

次に各月の平均値をこの總平均値に對する比率に改めればよい。これが季節指數であつて、その形は

$$\frac{x_1}{\bar{x}}\times100,\ \frac{x_2}{\bar{x}}\times100,\cdots\cdots\cdots\frac{x_{12}}{\bar{x}}\times100$$

である。指數の合計は一二〇〇になる。甚だ簡單であるが、二三の注意を加へておこう。

問題は月別平均のとり方である。七年間の資料を使へば七箇の一月の値、七箇の二月の値……七箇の十二月の値を、それぞれ平均するわけであるが、いま十月の値が次の如くであつたとする。

年次	十月の溫度
1	52
2	59
3	58
4	60
5	56
6	58
7	56

これから平均溫度を求めるには、(1)普通の算術平均か、(2)中位數か、(3)中央數項の算術平均か、この三者の何れかを使用するのが慣例である。(1)によれば57°、(2)によれば58°であるが、この差は第一年目の溫度が異常に低く（52°）、これが算術平均値を低からしめたゝめに生じたのである。われわれは正常値を求めてゐるのであるから、かような異常値は計算に入れない方がよい。そのためには(2)の如き中位數をとるか、又は異常値を除いた算術平均をとればよい。中位數については既に說明したが、後者は先づ最大値（60°）と最小値（52°）とを取除き、次に殘りから最大値（59°）と最小値（56°）を取除き、項が多い場合はこの手續きを何回も繰返へせば、結局中央數項が殘る。これは極端な卽ち異常な値を含まない項のみであるから、それを平均すれば最も安當な正常値が得られるわけである。前例について中央三項平均をとれば57.3°となる。氣象臺あたりで最も廣く用ひてゐるやうである。算術平均と中位數との折衷的方法であるが、その際注意を要することは、必ず兩極端の値を一對づつ取除いてゆくことで、前例でいへば、52°だ

けが特に異常だといつてそれだけ取除いてはいけない。

二　社會現象に於ける季節指數の算定

自然現象に適用されるこの方法は單純月別法である。かような現象に於ける異常値は全く偶然的原因に基くから、單に平均によつて除去されるのであるが、社會現象の殆ど全部は、偶然的原因のみでなく、他の原因による變化をも免れない。電力消費量は、勿論季節によつて甚だちがふが、同時に文化の進歩と共に增加してゆく傾がある。この發展的傾向を除去しなければ、正しい季節指數は求められないのであつて、それには、對傾向値比率法、十二ケ月移動平均法及び連環比率法の三つが最も廣く利用されてゐる。

(1)　對傾向値比率法

一九〇四年から一九一三年までのフランクフルト市の電力消費量について上記の單純法を適用すれば、次の結果を得られる。

	四月	五月	……	三月	總平均
平均	357.3	262.1	……	474.2	492.8
指數	72	53	……	96	100

ところがこの十年間の發展傾向は極めて顯著で、一九〇四年には月平均二五九・二であつたものが、一九一三年には七二〇・二、即ち三倍近く激增してゐる。故に單純法で求めた指數は、年度初には低く年度末には高くなつてゐるわけで、その分だけ修正されねばならない。いま與へられた資料について直線傾向線を當嵌めれば、一年間の增加分は四八・四となる。（$y=a+bx$ に於けるbの値である）。これを年平均 492.8 で割れば 0.0984 となり、從つて一ケ月の增加分はその十二分の一即ち 0.0082 である。四月に始まる一年間の上記指數は前半即四月から九月までは實際よりも低く、後半即ち十月から三月までは實際よりも高くなつてゐるから、この修正分によつて修正せねばならぬ。即ち中央の二ケ月（九月と十月）については

九月、$1-\frac{0.0082}{2}=0.9959$

十月、$1+\frac{0.0082}{2}=1.0041$

なる修正分を計算し、これでそれぞれの月平均を割ればよい。即ち

九月、$\frac{404}{0.9959}\fallingdotseq 406$

十月、$\frac{562}{1.0041}=560$

となる。八月の修正分は $0.9959-0.0082=0.9877$、十一月のそれは $1.0041+0.0082=1.0123$で

ある。かやうな計算を全部について行へば、上記の單純法の結果は次の如く修正される。

	四月	五月	六月	七月	八月	九月	十月	十一月	十二月	一月	二月	三月	平均
月平均	357	262	222	220	275	404	562	749	920	809	660	474	
修正分	0.955	0.963	0.971	0.979	0.988	0.996	1.004	1.012	1.021	1.029	1.037	1.045	
修正サレタ月平均	374	272	228	225	278	406	560	739	900	786	636	454	488
Trendヲ除去セル季節指數	77	56	47	46	57	83	115	152	185	161	130	93	

（註1） 幾分の不正確を覺悟すれば、最小自乘法によらずとも大體の增加分は算出される。即ち第一年目（一九〇四年）の平均（二五九・二）と最後の年（一九一三年）の平均（七二〇・二）との差を「年の數から一を除いたもの」即ち此處では九で割ればよい（$\frac{720.2-259.2}{9}=51.$）。年の數で割らずにそれよりも一だけ少い數で割る理由は、最初の年は全體の基準とされるからである。この五一と、最小自乘法による四八・八との差は極めて僅かであるから、この例に於ては斯かる省略法を用ひて差支へがないのである。

(2) 十二ケ月移動平均法

季節變動は一年間の變化であるから、一年間即ち十二ケ月を平均して了へば季節の影響は完全に抹殺される。十二ケ月移動平均法とはこの原理によつて原系列から季節變動を除去する方法である。Aから季節變動を除去したものがBと判れば、季節變動はA－Bまたは$\frac{A}{B}$として容易

に算出される。次の資料は一九一〇年一月から一九二一年十二月までのアメリカに於ける鷄卵ダース卸賣價格（仙）である。

鷄卵價格

	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921
一月	30.5	30.4	29.5	26.8	30.7	31.6	30.6	37.7	46.3	57.2	64.8	61.1
二月	23.9	22.1	29.1	22.8	28.4	29.2	26.8	35.8	49.4	48.3	56.9	49.6
三月	22.9	16.5	24.5	19.4	24.2	21.3	21.2	33.8	40.4	33.1	46.6	29.2
四月	18.6	14.9	17.8	16.4	17.6	16.6	17.9	25.9	31.2	34.3	38.8	20.4
五月	18.6	14.7	17.1	16.1	16.8	17.1	18.1	30.0	31.0	36.8	37.4	20.2
六月	18.3	14.5	16.7	16.9	17.3	16.6	19.0	31.1	29.8	38.6	37.0	19.4
七月	18.2	14.2	16.7	17.0	17.6	16.8	19.7	28.3	30.7	36.8	36.7	22.0
八月	17.6	15.5	17.4	17.2	18.2	17.0	20.7	29.8	34.4	39.3	40.0	26.6
九月	19.4	17.4	19.1	19.5	21.0	18.7	23.3	32.2	36.4	41.0	44.2	30.4
十月	22.4	20.0	22.0	23.4	23.5	22.3	28.1	37.4	41.6	44.7	50.1	34.2
十一月	25.3	23.5	25.9	27.4	25.3	26.3	32.2	39.4	47.2	54.0	56.9	44.2
十二月	29.0	28.7	29.7	23.0	29.7	30.6	38.1	43.3	55.0	61.9	65.0	51.1
平均	22.5	19.4	22.1	21.5	22.5	22.0	24.6	33.8	39.5	43.8	47.9	34.0

十二ケ月移動平均

	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921
一月	—	20.25	21.26	20.78	22.74	22.29	22.24	30.06	36.72	41.39	46.60	40.47
二月	—	19.99	21.45	20.79	22.80	22.20	22.51	30.80	37.01	41.86	46.63	39.30
三月	—	19.82	21.60	20.79	22.91	22.06	22.86	31.59	37.33	42.25	46.80	38.16
四月	—	19.64	21.75	20.87	22.97	21.91	23.29	32.39	37.64	42.57	47.15	36.93
五月	—	19.46	21.94	20.99	22.89	21.90	23.77	33.08	38.14	42.98	47.50	35.74
六月	—	19.38	22.08	21.19	22.66	21.98	24.33	33.59	38.96	43.56	47.75	34.63
七月	22.47	19.33	22.00	21.49	22.57	21.98	24.94	34.17	39.90	44.17	47.72	—
八月	22.18	19.58	21.63	21.88	22.64	21.84	25.60	35.10	40.31	44.84	47.26	—
九月	21.63	20.20	21.16	22.32	22.55	21.74	26.51	35.93	39.96	45.76	46.24	—
十月	21.21	20.66	20.89	22.57	22.39	21.78	27.36	36.43	39.79	46.50	44.74	—
十一月	20.89	20.88	20.79	22.65	22.36	21.88	28.20	36.69	40.17	46.71	43.26	—
十二月	20.57	21.07	20.76	22.69	22.35	22.02	29.20	36.68	40.77	46.67	41.81	—

先づ一九一〇年一月から同年十二月までの一年間を平均し、得た二二・四七仙を同年七月に置く。次に一ケ月ずらして二月から次の一九一一年一月までの一年間を平均し、得た二二・一八を八月に置く。かく順次一ケ月づゝずらして最後まで計算すれば次の表をうる。最初と最後の各六ケ月は、置かれる値はないわけである。いま季節變動を比率として求めれば次の表の通りとな

る。その値は原資料の各月の値を、その月の移動平均値で割つたものである。

	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921
一月	—	150.1	138.8	129.0	135.0	141.8	137.6	125.4	126.1	138.2	139.1	151.0
二月	—	110.6	135.7	109.7	124.6	131.5	119.1	116.2	133.5	115.4	122.0	126.2
三月	—	83.2	113.4	93.3	105.6	96.6	92.7	107.0	108.2	78.3	99.6	76.5
四月	—	75.9	81.8	78.6	76.6	75.8	76.9	80.0	82.9	80.6	82.3	55.2
五月	—	75.5	77.9	76.7	73.4	78.1	76.1	90.7	81.3	85.6	78.7	56.5
六月	—	74.8	75.6	79.8	76.3	75.5	78.1	92.6	76.5	88.6	77.5	56.0
七月	81.0	73.5	75.9	79.1	78.0	76.4	79.0	82.8	76.9	83.3	76.9	—
八月	79.4	79.2	80.4	78.6	80.4	77.8	80.9	84.9	85.3	87.6	84.6	—
九月	89.7	86.1	90.3	87.4	93.1	86.0	87.9	92.4	91.1	89.6	95.6	—
十月	105.6	96.8	105.3	103.7	105.0	102.4	102.7	102.7	104.5	96.1	112.0	—
十一月	121.1	112.5	124.6	121.0	113.1	120.2	114.2	107.4	117.5	115.6	131.5	—
十二月	141.0	136.2	143.1	145.4	132.9	139.0	130.5	118.1	134.9	132.6	155.5	—

この各月の比率は何れの年も大差ないが、なほ若干の不一致があるから、代表的正常値を求めねばならぬ。算術平均でも中央數項のそれでもよいが、中央數項をとれば次の如くである。その總平均九九・六で各々を割れば、最後の欄の季節指數が得られる。

月	中位數	指　數
1	138.2	138.7
2	122.0	122.5
3	96.6	97.0
4	78.6	78.9
5	77.9	78.2
6	76.5	76.8
7	78.0	78.3
8	80.4	80.7
9	89.7	90.1
10	103.7	104.1
11	117.5	118.0
12	136.2	136.7
平均	99.6	100.0

(3)　連環比率法

アメリカの著名な經濟統計學者パーソンズ（Persons）の案出した方法で、最も廣く利用されてゐる。季節變動は既に述べた通り月から月への變動であるから、與へられた月別統計を悉く直前の月に對する比率に換算し（これを連環比率 Link-relatives といふ）これから指數を求めんとするものである。次の資料はベルリン取引所市場割引率で、次表はその連環比率である。例へば一九〇〇年二月の比率〇・九五は、二月の値四・二一を直前の一月の値四・四二で割つたものである。

Berlin 取引所市場割引率

	一月	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月
1900	4.42	4.21	5.21	4.43	4.56	4.86	4.06	4.03	4.41	4.03	4.16	4.49
1901	3.57	3.22	3.79	3.37	3.19	3.20	2.81	2.26	2.68	2.83	2.84	2.96
1902	2.11	1.85	1.79	1.65	1.98	2.17	1.59	1.73	2.14	2.73	3.11	3.38
1903	2.26	1.90	2.69	2.61	3.09	3.29	2.96	3.30	3.68	3.32	3.46	3.54
1904	2.58	2.77	3.44	2.83	3.10	2.98	2.60	2.62	3.09	3.69	3.99	3.94
1905	2.56	1.93	2.22	1.91	2.30	2.34	2.12	2.23	2.99	4.00	4.62	4.99
1906	3.81	3.35	4.02	3.44	3.39	3.68	3.49	3.43	4.23	4.83	5.27	5.58
1907	4.90	4.68	5.40	4.65	4.44	4.66	4.44	4.62	5.08	4.91	6.61	7.07
1908	4.98	4.48	4.49	4.11	3.91	3.33	2.76	2.82	3.14	2.79	2.54	2.92
1909	2.24	2.17	2.66	1.98	2.32	2.91	2.28	2.13	3.06	3.83	4.47	4.34
1910	3.09	2.94	3.52	3.14	3.19	3.23	3.03	3.33	3.85	4.15	4.50	4.53
1911	3.50	3.07	3.34	2.96	2.84	3.38	2.46	3.03	4.16	4.32	4.51	4.86
1912	3.33	3.79	4.72	3.75	3.91	4.14	3.36	3.93	4.38	4.19	5.23	5.94
1913	4.68	5.15	5.90	4.56	5.31	5.53	4.65	4.88	5.35	4.71	4.45	4.57
1914	3.11											

	一月/十二月	二月/一月	三月/二月	四月/三月	五月/四月	六月/五月	七月/六月	八月/七月	九月/八月	十月/九月	十一月/十月	十二月/十一月
1900		0.95	1.24	0.85	1.03	1.07	0.84	0.99	1.09	0.91	1.03	1.08
1901	0.80	0.90	1.18	0.89	0.95	1.00	0.88	0.80	1.19	1.06	1.00	1.04
1902	0.71	0.88	0.97	0.92	1.20	1.10	0.73	1.09	1.24	1.28	1.14	1.09
1903	0.67	0.84	1.42	0.97	1.18	1.06	0.90	1.11	1.12	0.90	1.04	1.02
1904	0.73	1.07	1.24	0.82	1.10	0.96	0.87	1.01	1.18	1.19	1.08	0.99
1905	0.65	0.75	1.15	0.86	1.20	1.02	0.91	1.05	1.34	1.34	1.16	1.08
1906	0.76	0.88	1.20	0.86	0.99	1.09	0.95	0.98	1.23	1.14	1.09	1.06
1907	0.88	0.96	1.15	0.86	0.95	1.05	0.95	1.04	1.10	0.97	1.35	1.07
1908	0.70	0.90	1.00	0.92	0.95	0.85	0.83	1.02	1.11	0.89	0.91	1.15
1909	0.77	0.97	1.23	0.74	1.17	1.25	0.78	0.93	1.44	1.25	1.17	0.97
1910	0.77	0.95	1.20	0.89	1.02	1.01	0.94	1.10	1.16	1.08	1.08	1.01
1911	0.77	0.88	1.09	0.89	0.96	1.19	0.73	1.23	1.37	1.04	1.04	1.08
1912	0.69	1.14	1.25	0.79	1.04	1.06	0.81	1.17	1.11	0.96	1.25	1.14
1913	0.79	1.10	11.5	0.77	1.16	1.04	0.84	1.05	1.10	0.88	0.64	1.03
1914	0.68											

これから各月の代表値を求める。こゝでは中央八項の算術平均をとらう。連環比率法の特徴はこれから後の手續に現はれるのである。先づ連鎖比率(Chain relatives)を求める。それは一月の

値を直ちに一〇〇と置き、これを一月の連鎖比率とする。これに二月の平均値〇・九二を乘じた九二を二月の連鎖比率に、これに三月の平均値一・一九を乘じた一〇九を三月の連鎖比率とし、以下この計算を續けて、最後に十二月の連鎖比率一・三〇に一月の平均値〇・七三を乘じた九五を算出する。季節運動は一年で完全に一巡すべきであるから、最初の一月の一〇〇と、最後に算出した數字とは一致せねばならぬものである。然るにこの例では後者は九五で、一〇〇ではない。これは右十四年間の傾向が下降的なためで、もし逆に上昇的なら、一〇〇以上の數字となるのである。何れにしろその差の五は二月から最後の一月までの間に於て順次修正さるべき分である。これが最も簡單な修正方法はこの差を十二分し、得たる〇・四一六を二月の九二に加へ、その二倍の〇・八三二を三月の一〇九に加へ……その十二倍の五を最後の九五に加へればよい。斯くすれば一〇〇に始まつたものが一〇〇に終るわけで、これによつて季節變動の本質に合致した連鎖比率が得られる。最後にこの修正された連鎖比率の平均一〇七にて各比率を割れば

	中央八項の平均	連鎖比率
一　月	0.73	100
二　月	0.92	92
三　月	1.19	109
四　月	8.6	94
五　月	1.06	100
六　月	1.05	105
七　月	0.86	90
八　月	1·04	93
九　月	1.17	109
十　月	1.04	114
十一月	1.08	123
十二月	1.06	130
一　月		95

總和が一二〇〇となるところの季節指數が求められるのである。

	修正された連鎖比率	季節指數
一月	100	93
二月	92	86
三月	110	103
四月	95	88
五月	102	95
六月	107	99
七月	93	87
八月	96	90
九月	112	104
十月	118	110
十一月	127	118
十二月	135	127

併し最初の一月と最後の一月の値の差が算術的に十二ケ月に均分される修正法は理論的には缺陷がある。差が僅少なる場合にはこの方法で充分であるが、大なる差の生じた場合には寧ろ複利的に配分するのが正しいであらう。今、一月から次の一月までの連鎖比率を 100, C_2, C_3……C_{12}, C_{13} とすれば

$$100(1+d)^{12}=C_{13}=95$$

dが計算されゝば、修正された各月の連鎖比率は

$$100,\ \frac{C_2}{(1+d)},\ \frac{C_3}{(1+d)^2},\ \cdots\cdots\ \frac{C_{12}}{(1+d)^{11}},\ \frac{C_{13}}{(1+d)^{12}}$$

實際にこれを計算するには當然對數に書改めねばならぬ。次表はその手續きを示したもので、例へば三月の修正連鎖比率 $\frac{C_3}{(1+d)^2}$ は、對數では $\log C_3-2\log(1+d)$ となるから、一般に $\log C-(n-1)\log(1+d)$ によつて表はされる事は說明する迄もないであらう。

	C	logC	(n−1)×log(1+d)	logC−(n−1)×log(1+d)	修正された連鎖比率	指数
一月	100	2.00000	0.00000	2.00000	100	93
二月	92	1.96379	0.99814−1	1.96565	92	86
三月	109	2.03743	0.99628−1	2.04115	110	103
四月	94	1.97313	0.99442−1	1.97871	95	88
五月	100	2.00000	0.99256−1	2.00744	101	94
六月	105	2.02119	0.99070−1	2.03049	107	99
七月	90	1.95424	0.98884−1	1.96540	92	86
八月	93	1.96848	0.98698−1	1.98150	96	90
九月	109	2 03743	0.98512−1	2.05231	113	105
十月	114	2.05690	0.98326−1	2.07364	118	110
十一月	123	2.80991	0.98140−1	2.10851	128	119
十二月	130	2.11394	0.97954−1	2.13440	136	127
一月	95	1.97772	0.97772−1	2.00000		

平均＝107

三　短期の季節變動

右に述べたのは何れも一年を週期とする季節變動であつた。併しわれわれは、より短期の季節

變動を考へることもできる。一日二十四時內の變化はその一つである。人間の生活は每日可成り規則的で、一定の時間に起床し、食事し、勞働し、休息する。百貨店や電車には定まつたラッシュアワーがある。故に時間別の統計からその樣態を知りうるわけで、一種の季節指數といへよう。また學生々活や映畫館の入場者數は週の曜日によつて規制される。これらは日別統計から知りうることで、これも亦一種の季節變動である。百貨店、映畫館、電鐵會社等々は、かゝる變動の正確な知識の上にのみ適宜な營業政策を樹てうるのであつて、よく入口に計器をもつた調查員が頑張つてゐることは、諸君或ひは御存知ないかも知れない。電鐵會社になると、ときどき交通調查票を配布して公然と調查するから、これは誰でも知つてゐる。某百貨店ではこの調查に基いて店員の休暇日や休息時間なぞの割振りまで決めてゐる。

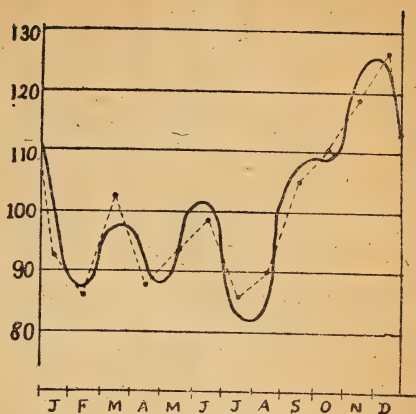
給料が週拂制だと、支拂日には大衆酒場が繁盛し、翌日は缺勤者が殖える。學生は月曜の朝の時間が一番出席率が惡い。日曜に遊び過ぎるからであらう。こうしたことから風紀や犯罪なぞにも或る規則性が生れるとのことである。勿論本來の意味に於ける季節がこれらに大きな影響をもつてゐることは、久しい以前から知られてゐる。春先から性犯罪や自殺が增加し、年末には經濟的犯罪が增加することは、理窟の上からでも納得できよう。近代統計學の先驅をなしたケトレ

ーの「人間に就いて」と題する書物には、これに關する多くの興味深い記述が見出される。例へば生命や身體に對する犯罪は夏に多く、財産に對するそれは冬に多いといつた類である。しかし右に述べたように、一つの季節の中に周期のより短い別の起伏もあるから、眞の、廣義の、季節變動の樣態は、與へられた資料を詳細に分析して見なければ判らないのである。その科學的な方法は入門書の範圍を遙かに超えてゐるが、次の一節によつてその構想を想像されたい。

季節變動は上記の如く指數で示され、例へばA商品の價格のそれは四月は七〇、五月は八〇といひあらはす。しかし一考すれば、かような表現は甚だ不充分なことが判る。それは各月の平均を言つてゐるに止まり、實際には決して四月中は七〇の線に固定し五月になると同時に一躍八〇になるといふものではない。季節は除々にそして連續的に推移するのであつて、從つて季節變動の眞相は矢張り連續的な形で示されねばならない。かくて科學的最密を要する研究に於ては、季節指數によつて描かれたグラフを三角函數を用ひて曲線化する方法が行はれてきた。こゝでは省略せねばならぬが、元來三角函數は週期運動を示すもので、季節變動を現はすには最適なのである。例へば前揭のベルリン取引所市場割引率の季節指數には

$$y=100+12.14\sin(117.5^\circ+x)+9.27\sin(150^\circ+2x)+8.68\sin(101.5^\circ+4x)$$

なるフーリェ曲線が當嵌められる。次圖の点線は季節指數を連ねたもの、實線はこの曲線を示す。詳細は P. Lorenz. Die Bestimmungsgründe für die Saisonschwankungen des Berliner Marktdiskonts in der Vorkriegszeit. (Vierteljahrshefte zur Konjunkturforschung, 1928, Heft4. TeilA) を見られたい。なほロレンツはこの研究に於て、季節を自然的及び人爲的の二つに分解する方法を展開してゐる。市場割引率の變化は收穫期といふ自然的季節により、また支拂方法の慣習といふ人爲的季節によつて大きな影響を受ける。この二つの原因がそれぞれいかなる影響を與へるか、これは算出された季節指數を更に分析することによつてのみ知りうることなのである。(第二十圖)

統制經濟の下では價格の季節變動は微弱とならざるを得ない。併し闇價格や從つて實效價格には明かに存在し、また統制緩和と共に次第に多くの商品が自由價格となつて、往時の季節變動を再現せんとしてゐる。人の生活が自然の規絆を脫し得ない以上、季節變動はいつ迄もわれわれの生活につき纒ふであらう

第二十圖

四　景氣變動と不規則變動

一つの與へられた時系列に於ける變化(α)は既述の如く發展傾向(T)季節變動(S)景氣變動(C)及び不規則變動(I)の四つの運動の綜合的結果である。この四つがいかなる形で結合してゐるかは素より言明し得ない。即ち

$$\alpha = f(T, S, C, I)$$

と言ひうるに過ぎない。併し結合の仕方について何等かの假定を設ければ、形式的にはαをその構成運動に分解しうることは勿論であつて、パーソンズを中心とするアメリカの景氣研究者はこれによつて景氣變動を抽出せんとした。いま或る時點t_1の原資料をy_1とする。t_1に於ける長期的正常値はt_1に於ける傾向値T_1であり、t_1に於ける季節的正常値はt_1の屬する月の季節指數s_1であるから、t_1時點の長期的且つ季節的正常値は$T_1 \times S_1$と考へることができよう。よつて

$y_1-(T_1 \times S_1)$ は該時點に於ける景氣變動と不規則變動の合成分Bを示すものであり、或ひはこれを比率化した

$$\beta=\frac{y_1}{(T_1 \times S_1)}$$

はその相對的大さを示すものである。例へば $y_1=230$圓, $T_1=210$圓, $S_1=90\%$ とすれば、

$$\beta=\frac{230}{210 \times 90\%}=121.6\%$$

となる。これは原資料から發展傾向と季節變動を除去した値で、景氣變動と不規則變動のみから成るものである。これを全時點について計算することによつて、全期間に亘るこの合成分の變化を知ることができる。問題はこの合成分をいかにして更に分解するかであるが、これについては充分滿足な方法はない。蓋し不規則變動は元々その原因が一定の規則をもたないからである。そこで最も無難な方法として、移動平均法が考へられる。これによつて不規則變動は可成りの程度に除去され、景氣變動が略々正確に摑めるであらう。斯くして得られた景氣變動指數が景氣豫則や景氣週期の研究に對する出發點となるのである。

第十三章　相關計算（一）

一　因果關係

物を正確に知るといふことが人の知的努力の普遍的目標であること、そして數字がその不可缺の手段であることは、最初に述べた通りである。對象が集團的事實であれば、これを正確に記述するものは統計といふ數字である。即ちわれわれは統計によつて集團的事實を正確に知ることができるわけである。併しこゝでわれわれは「正確に知る」とは一體何であるか、改めて考へて見る必要がある。物價騰貴といふ事實は物價指數によつて數字的に即ち正確に知ることができる。だがそれだけ知つて、一體何の役にたとう。いかに正確な知識も、それ自身獨立してゐては、無用の長物である。勿論それは人をビックリさせることはできよう。世の中にはそれが樂しくて無闇に物を覺えようとする物好きがゐる。ラジオの「話の泉」に出演すれば喝采を博することは確かだが、殘念ながらそれだけのこと、謂はゞ頭腦の遊戯に過ぎない。折角の努力が遊戯に終つて、何等の實際的效果も產まないとすれば意味のない話である。

知識はわれわれの合理的行動の指針となつて始めて意義を生ずる。物價騰貴といふ事實に關する知識は、單に騰貴の事實だけでなく、何故騰貴したか、そして騰貴すればどうなるかといふ原因及び結果を知ることによつて始めて完全な知識となり、始めてわれわれの行動の指針となるのである。因果關係は別の言葉でいへば理論であり、また法則である。物體に壓力を加へれば體積は縮小する。この因果關係をボイルの法則といふ。價格が變化すれば需要は反對の方向に變化する。この因果關係を需要の法則といふ。自然現象間の因果關係を自然法則といひ、これを記述し說明するのが自然科學の任務である。社會現象間の因果關係は社會法則であつて、これを記述し說明するのは社會科學の義務である。因果關係を認識することを一般に理論または法則を樹てるといふ。法則樹立が、事實を對象とする凡ゆる科學の使命と言はれるのはこれが爲である。して見れば統計學も物價騰貴の事實を記述するに止まらず、その原因及び結果の說明に役立つて、始めて眞の意義を獲得するのである。

併しこの問題でわれわれは統計學が常に一箇の形式科學であることを想起せねばならない。事物の因果關係を說明するのは常に實體科學の任務である。經濟學は經濟現象の因果關係を說明し、以て經濟法則を樹立する。物價騰貴の原因または結果を說明するのは、斯くて經濟學の任務

である。ところがこの同じ問題が統計學によつても取扱はれるとすれば、經濟學と統計學との差別は甚だ怪しくなる。併し統計學の形式科學としての性格を忘れなければ、この疑問は起らないで濟む。

既に述べたように統計學とは集團的事實を數字的に把捉する方法の學である。方法とは手段のことで、手段とは或る主體が或る目的のために使用するものである。鋸がなければ家は建たない。だが家を建てるのは大工であつて鋸ではない。正確な資料がなければ經濟學は成立しない。この正確な資料を提供するものは統計學であるが、それを組立てるものは經濟學である。通貨と物價の資料が與へられゝば、經濟學は例へば貨幣數量説といつた理論をたてることができる。そして一度びこの理論が提供されゝば、統計學はこれを檢討するためには活動する。併し統計學の範圍は資料の提供と理論の吟味といふ事前的及び事後的作用に限定されるのであつて、理論の樹立そのものは、統計を利用する實體科學の職務なのである。

ところが因果關係の決定は社會科學に於ては特に困難である。一つの原因から如何なる結果が生れるかは、他の要素を一定にし、一つの要素だけを變化させて見れば最も簡單に判る。これは物理的現象については實驗室で屢々行はれることで、一般に自然科學が多くの正確な理論をもつ

ゐるのはこのお蔭である。ところが社會科學では對象が社會的即ち集團的事實である關係から、實驗は一般に不可能である。そこでは種々雜多な原因と結果が錯綜し、それらが不斷に變化してゐて、一つの要素のみを變化せしめるといふことは望み難いのである。斯くて一つの事實について、いくつもの理論が生れるのである。價格低落の原因を、或ひは生産過剩に求め、或ひは消費不足に求め、以て生産過剩說と消費不足說とが對立することになる。何人をも承服せしめるような理論が社會科學で容易に求め難いのはこのためであるが、この故にこそ統計的方法による吟味が必要なのである。勿論このことは大きな理論についてばかりではない。物理的現象でも、例へば或る新藥が有效かどうかは投藥の結果から判定する外はないが、それには多數の患者について平均的結果を求めねばならぬ。況や價格騰貴がどれだけ需要を減退させるかは、廣範圍の觀察から始めて言へることである。

統計的方法によれば關係の厚薄が數字的に把捉される。供給量が減れば價格が上るといふことは凡ゆる商品について一樣に當嵌まるが、その程度は極めて區々である。勿論この場合にも、必需品に於ては需要は大して減らないから價格は暴騰する、即ち彈力性は小であり、奢侈品についてはその逆で、彈力性は大であるといつた區別なら、必ずしも統計によらずともできよう。併し正

確な知識は數字的でなければならぬといふ原則は、この場合にも當嵌まるのであつて、關係に厚薄があれば、その程度は矢張り數字的に示されることが望ましい。前例でいへば、供給量と價格との二つの統計から數字的に彈力性係數を決定するといふことである。これに應ずる統計的計算を相關計算といふ。

二　相關々係

右に述べた通り統計學は因果關係の決定は問題にしない。他の學問がこの關係ありと認めて提出した事象間の關係の程度を測ればよいのである。その場合の資料は統計といふ數字であるから、要するに數量間の關係を數字的に表現するにはどうすればよいかといふことに歸着する。ところが數學では數量間の關係を一般に函數關係といふ。いまx及びyなる二つの變數があつて、xが一定の値をとればyの値が決定される場合、yはxの函數だといひ、これを$y=f(x)$と記す。xを自變數、yを從屬變數といふ。ところが$y=f(x)$なら、$x=f(y)$も成立つ。換言すれば、xとyとは何れを自變數、何れを從屬變數としてもよい。これは數學も亦純然たる形式科學で、因果關係を問題にしないからである。即ち函數關係は二變數の相互的關係であつて、これを

共變關係と見ることができる。ところがいま價格Pと需要量Dの二つの系列を併記して見ると、Pが大きいときはDは小さく、Pが小さいときはDは大きいことが判る。然らばこれは矢張り共變關係であつて、而もこの場合は D＝f(P) なる函數關係が成り立つように見られる。併し函數といふ文字をその本來の數學的意味に解釋すれば、需要と價格との間には斯かる關係は成り立たない。蓋しそれが成り立つためには、例外的事例はあつてはならないからである。需要法則なるものは、一つの傾向を言つてゐるに過ぎないのであつて、箇々の事例に於ては、價格が可成り上つても格別需要を差控へない人があつたり少し上れば全然差控へて了ふ人もあり、要するに如何なる場合にも妥當する法則ではないのである。經濟學の諸々の法則は實は全部がこの性質のもので、この點から經濟學では法則の文字は寧ろこれを傾向といふ文字で置換へた方がよいといふ人もある。何れにしろ經濟法則は、普遍妥當の自然法則とちがつて、單に多數事例から平均的に言ひうるところの統計的法則に外ならない。よつてわれわれは斯かる共變關係を函數關係から區別するために相關々係と名づける。それは謂はゞ弛緩せる函數關係である。ではこの關係はどうして測定できるか。

二つの統計集團の關係は言ふ迄もなくこれを平均的に規定する外はない。併し一度び平均的に

規定されゝば、複雜は單純化されて、本質が現はれる。別の言葉でいへば、例外的事例は總べて相殺されて、單純な函數關係に變形される筈である。斯くてわれわれはD=f(P)と規定することができるのであつて、それは科學に共通な單純化或ひは理論化の一典型であり、これを以て事實の歪曲と見るのは當らないのである。複雜なものを複雜のまゝに觀察し理解することは人の能力の及ばざることで、我々は枝葉を棄て末節を抽象することによつてのみ事物の本質に迫ることが出來る。このことは平均値の本質に關聯して既に述べたつもりである。即ち經濟的事實の本質を把捉し說明するのが理論經濟學の目的たる以上、單純化の過程によつて複雜な相關々係を簡單な函數關係に引き直すことは、單に差支へないばかりでなく、必要でもあるのである。數理經濟學の學問的性格と偉力とは實にこゝに在ると言へるであらう。

併し乍ら單純化或ひは理論化も、その目的は結局は複雜なる現實の理解に在る以上、いつ迄も自己滿足に留まることは許されない。それはどこ迄も現實理解の手段であつて、これを忘却すれば學問は一種の遊戲に終つて了ふ。我々は經濟量の間の共變關係を、最初には函數關係と假定して定式化するが、次の段階に於てはこれを武器として有りのまゝの現實の世界に直面せねばならぬ。別の言葉で言へば相關々係の把捉に移行せねばならぬ。このことは理論の妥當性を決定する

手段としても極めて必要なることである。蓋し函數化によつて成立した理論は、現實と極めて接近してゐる場合もあるし、さうでない場合もありうるから、何等かの手段によつて、理論と現實との距離が規定されざる限り、事物の理論化は無制限に押進められ、虚空に遊離した蜃氣樓となる惧れがあるからである。

函數の取扱ひが數學の問題たるに對して、相關のそれは統計學の課題である。蓋しそれは現實の複雜性そのものから出發せねばならぬからである。併し乍らその際にも猶ほ且つ尠からざる抽象化は避け難い。第一に、相關の測定は現實から出發するとはいへ、無限に錯綜せる複雜をそのまゝ出發點とすることは元々不可能で、結局我々は極めて狹小なる範圍の相關、即ち一變數と他の一變數乃至二三の變數との間の相關に限定せざるを得ないのである。大風が吹けば砂塵が卷き上るから盲人が殖え、三味線の需要が殖えて猫が少くなり、鼠が殖えるから桶が嚙られて桶屋が繁盛するといふ複雜な關係は、必ずしも落語の笑話ではなく、世の中の總べてが斯かる廻り廻つた因緣に結ばれてゐるのである。相關の問題でこれらの諸關係を一々採り上げねばならぬとすれば、全くやり切れたものではない。故に前例を以てすれば、盲人の增加と三味線に對する需要增加の關係とか、鼠の增殖と桶屋の利潤との關係といふが如き、謂はゞ部分的問題に限定するので

ある。

第二には、相關々係は函數關係の複雜化したものであるから、その基本形態は依然函數關係であり、從つて相關測定にはこの基本形態が基準とされるといふことである。次に述べるが如く、相關測定は回歸線なるものを基準として行はれるが、回歸線とは二變數間の函數關係を示す數學式に外ならない。故に相關理論は函數理論と對立するものではなく、寧ろその擴充と見做さるべきものなのである。

相關理論は統計學の最も困難なる問題の一つであるから、こゝにその詳細を盡すことは不可能であるが、以下その要點たる回歸線、標準誤差、相關係數なぞについて、その一般を略述しよう。

三　回　歸　線

函數關係でない共變關係は如何にして函數關係に引直されるか。このことは數學線の當嵌に於て既に說明したことである。xとyとが函數關係にあるときは、xの値とyのそれとは一定の規則に從つて配列されるから、それらを座標點として描けば一定の數學線になる。これに反してxとyとが函數關係になければ、座標點は不規則に散在し、從つてそれらを連ねたジクザックの不

規則な形になる。最小自乘法に從つてこれらの散在的座標點の間へ數學線を引けば、この線を示す方程式は、xとyとが函數關係に立つものと假定して其の形を表したものである。xが時の經過を示す場合は、この線を傾向線といひ、其他の場合はこれを回歸線又は退行線といふ。相關の測定で問題となるのは事物と事物との關係であるから、引かれた線は傾向線ではなくて回歸線である。計算方法は傾向線の場合と同一であるから、一例を擧げるに止める。表の一番左の1.2.3.……は年次、xは物價指數、yは賃銀指數である。

年次	x	y	xy	x^2
1	158	101	15958	24964
2	146	96	14016	21316
3	141	96	13536	19881
4	134	95	12730	17956
5	130	95	12350	16900
6	137	100	13700	18769
7	150	103	15450	22500
8	158	104	16432	24964
9	153	104	15912	23409
10	145	105	15225	21025
計	1452	999	145309	211684

傾向線の場合の公式即ち

$$\begin{cases} \Sigma(y) = Na + b\Sigma(x) \\ \Sigma(xy) = a\Sigma(x) + b\Sigma(x^2) \end{cases}$$

をそのまゝ利用すればよい。よつて表の上で$\Sigma(y)$, $\Sigma(xy)$, $\Sigma(x^2)$ を計算して公式に代入すれば次の如くである。

$$\begin{cases} 999 = 10a + 1452b \\ 145309 = 1452a + 211684 \end{cases}$$

$\therefore\ a=50.7 \quad b=0.297$

故に求める回歸線の方程式は

$$y=50.7+0.297x$$

となる。これは右資料に於ける物價と賃銀との最も一般的な關係である。或る物價に對應する最も確からしい賃銀は、この方程式のxにその物價を代入することによつて求められる。逆に、或る賃銀に對する最も確からしい物價も同様にして求められる。これには右方程式のxとyとを入替へ、$x=a+by$ を求めればよいから、計算は次のようになる。(X^2 の代りに Y^2 を求めることに注意されたい。)

年次	X	Y	XY	Y²
1	158	101	15958	10201
2	146	96	14016	9216
3	141	96	13536	9216
4	134	95	12730	9025
5	130	95	12350	9025
6	137	100	13700	10000
7	150	103	15450	10609
8	158	104	16432	10816
9	153	104	15912	10816
10	145	105	15225	11025
計	999	1452	145309	99949

$$\therefore \begin{cases} 1452 = 10a + 999b \\ 145309 = 999a + 99949b \end{cases}$$

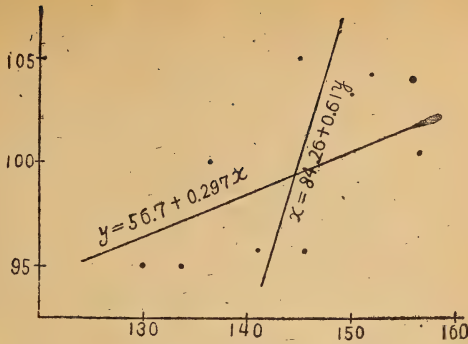
$$\therefore \quad x = 84.26 + 0.61y$$

ところが右に得た二つの方程式の結果は一致してゐないことに留意されたい。物價一三〇に對應する賃銀の理論値は第一方程式のｘに一三〇を代入したときのｙの値即ち

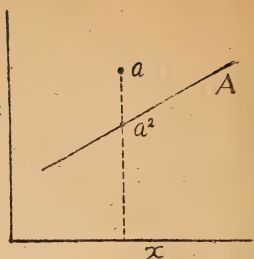
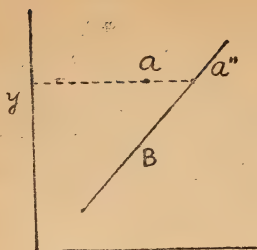
$$y = 56.7 + 0.297 \times 130 = 95.3$$

である。然るに第二方程式から九五・三の賃銀に對應する物價の理論値を求めれば

$$x = 84.26 + 0.61 \times 95.3 = 142.4$$

となつて、一三〇とはならない。即ち物價を自變數とするか、賃銀を自變數とするかによつて、全く異る回歸線が得られるのであつて、これを圖示すれば次の如くである。（第二十一圖）

第二十一圖

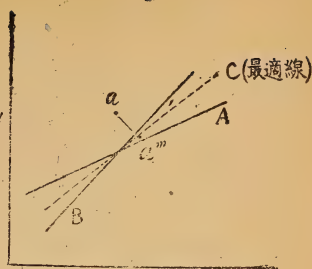
二變數間の相關計算に於て、その何れを自變數とするも隨意ではあるが、結果がかく異る以上、實際問題としてはその選擇は甚だ重大と言はねばならない。一體かく結果が異るのは何故かといへば、自變數として選ばれたものは基準であり、當然正しいもの即ち誤差を伴はざるものと假定されるに對し、從屬變數として選ばれたものは偶然誤差を免れないと想定されてゐるからである。前例で言へば、もしx軸に物價をとれば、物價統計は正確で、これに對する賃銀統計は偶然誤差を伴ふと假定されたのである。圖Aの座標點aは、賃銀のみの誤差によつて回歸線から離れたものとすれば、正しい位置はa′でなければならぬ。（第二十二圖）

次にもし賃銀を自變數とすれば、誤差は物價側から起つたことゝなり、圖Bに於けるが如く、a

の正しい位置はa″でなければならない（第二十三圖）。即ち何れを自變數とするかによつて、正常點の所在は異るわけで、從つて正常點の軌跡と考へられる回歸線も異らざるを得ないのである。時系列について傾向線を求める場合には、時を自變數として何等の不合理も起らない。蓋し時の經過は規則的で、誤差は伴はないからである。然るに回歸線に於ては何れを自變數とするかは、何れの誤差がより少いかによつて決める外はないが、それは多くの場合不可能にちかい。最初に一方を、次に他方を自變數として二つの回歸線を描いて見て、兩者が比較的一致すれば、二つの資料の正確度は同程度と判定され、どちらを自變數としても差支へないことになる。兩者が甚だしく異れば、どちらの回歸線も信用できないわけで、別の考慮が必要となるのである。

第二十二圖

第二十三圖

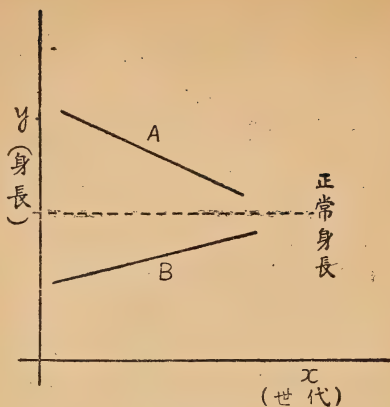
その場合、誤差は双方の變數に含まれると考へられるから、正しい回歸線は、右の二つの回歸線

を綜合した一種の平均線でなければならない。これを最適線(Line of best fit)または相互回歸線(Line of mutual regression)といふ。(第二十四圖)

その說明は初等統計學の範圍でないが、要するに座標點からx軸またはy軸に垂直に測つた偏差の平方の和を極小ならしめる代りに、回歸線そのものへ垂直に測つた偏差の平方の和を極小ならしめる手段を講ずればよいのである。この方法には或る種の非現實的假定が含まれて居り、且つ今までのところ曲線回歸線には適用されてゐない。この方法の利用例については

第二十四圖

Henry Schultz, Statistical Laws of Demand and Supply, 1928 を參照されたい。

回歸又は退行といふ文字は聊か說明を要する。始めて斯かる線を使用したのは英國の生物學者ゴルトン(F. Galton)である。彼は親子の身長の間には密接な關係があつて、長身の親からは長身の子供が、短身の親からは短身の子供が生れる傾きあること、即ち正の相關々係のあることを證明したが、同時にその相關度は世代の經過と共に次第に薄らいでゆき、子供は次第に普通の身長に近づいてゆくといふ事實を發見したのである。故に同一血統の親子を數代に亙つて觀察すれ

ば、異常に長身の先祖から出た家系では次圖のA線の如く、異常に短身の先祖から出たそれではB線の如く、何れも普通の身長に向ふのであつて、斯く異常値は次第に正常値に復歸するといふ意味で回歸又は退行なる文字を使用したのである（第二十五圖）。今日では回歸線又は退行線は單に二つの相關的變數を函數關係として表はした線を意味し、回歸とか退行とかの意味は全く含まれてゐない。即ちそれは單なる歷史的名稱に過ぎないのである。

第二十五圖

回歸線は必ずしも直線とは限らない。嵌線法の箇所で述べた通り、二變數間に拋物線其他の曲線も當嵌めうるのであつて、從つて種々なる形の曲線的回歸線がありうるのである。與へられた資料に直線を當嵌むべきか曲線を當嵌むべきかは一に資料の性質によるのであつて、その原則は嵌線法の箇所で述べておいた。併し相關論で最も屢々現はれるのは直線的回歸線である。そしてこれから導かれた相關度測定の係數が後に說明する相關係數で、最も廣く利用されるものである。

また相關には順相關と逆相關の區別がある。一方の增大が他方の增大に、一方の減少が他方の減少に對應する正比的關係を順相關といひ、一方の增大が他方の減少を、一方の減少が他方の增大に對應する反比的關係を逆相關といふ。通貨量と物價、能率と賃銀、價格と供給量……の間には順相關が、資本量と利率、價格と需要量……の間には逆相關が原則である。順相關は右上りの、逆相關は右下りの回歸線によつて示される。但し順または逆を言へない場合は極めて多い。年齢と婚姻の關係を見るに、女子の場合は二十四五歲までは急カーブで上昇し、以後は徐々に低下してゐる。即ち最初の部分は順で、後の部分は逆である。年齢と死亡率の場合には最初の二三年間は逆で、それ以後は順となる。これらの場合は回歸線は必らず極大點又は極小點をもつ曲線となり、その點を境として順逆が區分されるから、これを別々に取扱はない限り、順逆を言ふことはできないのである。二次以上の曲線は一般に極大極小をもつから、かような線で示された相關々係は順逆を區別し難い。直線回歸線の場合は上昇または下降がはつきりしてゐるから、プラス記號によつて順を、マイナス記號によつて逆を示すことになつてゐる。

第十四章　相關計算（二）

一　標準誤差

相關關係を強いて函數の形に引直したものが回歸線であるから、回歸線は多かれ少かれ現實を歪曲した抽象物である。そしてその抽象性は場合によつて著しく異る。回歸線に變形されたことによつて現實の關係が過度に歪曲されるならば、斯かる回歸線は純然たる理論的構造物たるに過ぎない。統計的方法は、元來事物の眞相を數字的に明確にする爲のものであるから、斯くの如き抽象物を造出して滿足することは出來ない。即ち回歸線によつて變數間の一般的關係を示したならば、次にはその際行はれた抽象の程度を數字的に算出し、以て現實との距離を明示する必要があるのであつて、この目的に應ずるものが標準誤差である。

標準誤差は平均値の標準偏差と類似した概念である。第五章で述べた通り、平均値は與へられた一數列を代表する一箇の數字であるから、單なる抽象値に過ぎない。然るにその抽象度は標準偏差 σ によつて數字的に示されるのであつて、算術平均値（M）と各項 a_1 a_2 …… a_n との偏差（d）

の平方の算術平均の平方根、即ち

$$\sigma=\sqrt{\frac{(M-a_1)^2+(M-a_2)^2+\cdots+(M-a_n)^2}{N}}=\sqrt{\frac{M(d^2)}{N}}$$

なることは既に說明した。回歸線の抽象度も殆ど同じ方法で決定しうるのであつて、異る點は測定基準が、標準偏差の場合には平均値という一箇の値なるに反し、標準誤差では回歸線という一箇の線、即ち點の連續或ひは軌跡なることである。前例によつて言へば、物價xが一五八なるときは、賃銀yは現實では一〇一であるが、理論的には（即ち回歸線方程式のxを一五八として計算したyの値は）一〇三・六となり、二・六の差がある。同樣に次年度の物價一四六に對應する賃銀は、現實には九六、理論では一〇〇・〇で、その差は四・〇となる。斯くの如くxに對應するyの値につき一々現實値と理論値との差を求め、これらを前條の偏差と同樣に取扱へばよいのである。但しこの差は、偏差と區別するために觀測誤差又は單に誤差とづける。賃銀の現實値 y_1 y_2 ……y^n に對應するその理論値を Y_1 Y_2 …… Y_n で示せば、標準誤差Sは

$$S=\sqrt{\frac{(y_1-Y_1)^2+(y_2-Y_2)^2\cdots\cdots+(y_n-Y_n)^2}{N}}=\sqrt{\frac{M(y-Y)^2}{N}}=\sqrt{\frac{M(d^2)}{N}}$$

となる。前例について計算すれば次の如くである。

	x	y	Y	d	d^2
1	158	101	103.6	−2.6	6.70
2	146	96	100.0	−4.0	16.00
3	141	96	98.5	−2.5	6.20
4	134	95	94.9	+0.1	0.01
5	130	95	95.3	−0.3	0.09
6	137	100	97.4	+2.6	6.00
7	150	103	101.2	+2.8	7.80
8	158	104	103.6	+0.4	0.16
9	153	104	102.1	+1.9	3.60
10	145	105	99.7	+5.3	28.00
					74.53

$$S=\sqrt{\frac{\Sigma(d^2)}{N}}=\sqrt{\frac{74.56}{10}}=2.7$$

二　相關比と相關係數

回歸線の標準誤差は直ちにこれを相關度測定に利用することが出來る。蓋し相關度は標準誤差と反比例するからである。併し乍ら恰も標準偏差と同樣、それは絕對數であり且つ有名數である

から、一般的用途にはこれを比率に換算する必要がある。標準偏差に於ては、それを算術平均値に對する比率、即ち變化係數に改めたが、標準誤差に於ては標準偏差に對する比率即ち S_y/σ_y とする。蓋しもし現實値が眞に函數關係を示すときは、即ち最も強度の相關關係を示すときは、標準偏差の如何に拘らず標準誤差は零、從つて $S/\sigma=0$ となり、また全く相關關係の缺如するときは、標準誤差は標準偏差に一致し、從つて $S_y/\sigma_y=1$ となるべく、普通の關係は斯くて零から一までの間の數字によつて示されるからである。實際の計算では右の比率を平方し、且つこれを一から引いて

$$n^2=1-(S_y/\sigma_y)^2$$

即ち

$$n=\sqrt{1-(S_y/\sigma_y)^2}$$

とする。これによれば完全な相關關係即ち函數關係は一、完全な無關係は零、普通の關係は零から一までの間の數字によつて示される。このnを相關比（Correlation Ratio）といふのである。そして直線相關即ち直線回歸線を以て示される相關に於ては、この相關比を特に**相關係數**（Coefficient of Correlation）といひ、一般にrなる文字を以てこれを示す。且つその場合では

式そのものも次の如く變形される。

$$r=\frac{Mxy}{n\sigma_x\sigma_y}=\frac{Mxy}{\sqrt{Mx^2My^2}}$$

nは項（相對する一組、即ち相關圖の點）の數、xはX系列の各項の値と平均値との偏差、yはY系列の各項の値と平均値との偏差、σ_xはX系列の標準偏差、σ_yはY系列のそれを意味する。前例について計算すれば次の通りである。

年次	X	Y	x	y	xy	x^2	y^2
1	158	101	+13	+1	+13	169	1
2	146	96	+1	−4	−4	1	16
3	141	96	−4	−4	+16	16	16
4	134	95	−11	−5	+55	121	25
5	130	95	−15	−5	+75	225	25
6	137	100	−8	0	0	64	0
7	150	103	+5	+3	+15	25	9
8	158	104	+13	+4	+52	169	12
9	153	104	+8	+4	+32	64	12
10	145	105	0	+5	0	0	25
	平均＝145	平均＝100			+254	854	149

$$r=\frac{\Sigma xy}{n\sigma_x\sigma_y}=\frac{254}{10\sqrt{\frac{854}{10}\times\frac{149}{10}}}=+0.71$$

$$r=\frac{\Sigma xy}{\sqrt{\Sigma x^2\Sigma y^2}}=\frac{254}{\sqrt{854\times149}}=+0.71$$

相關係數が零から1までの大さをとることは判つたが、この例に於ける0.71が果してどれだけの關係強度を意味するかは困難な問題である。併し大雜把にいへば、零から0.2乃至0.3なら關係は殆ど問題にならず、0.5前後なら幾分關係があり、0.7前後なら可成り密接といへるし、0.8以上なら著しく強い關係があると見てよからう。なほ係數の前に附される正負の記號によつて關係の順か逆かを判定すべきことは勿論である。

註　完全な無相關に於ては回歸線はYの中央を通つてX軸に平行な直線となる。故に回歸線の上下にとつた3Sと系列Yに於ける平均値の上下にとつた3σとは、その幅全く等しい。即ちSとσとは一致する。後章の誤差理論を參照されたい。

三　多元相關と部分相關

いま迄に述べたことは何れも二變數間の相關即ち單純相關に關するものである。併し既に述べた通り、一事象を決定する要因は一般に複雜である。既に函數關係を語るに際し、クールノーが一商品に對する需要をその商品の價格の函數即ち $D_a = f(p_a)$ と見たのは不備であり、寧ろ幾多商品の價格函數即ち

$$D_a = f(P_a, P_b, P_c \ldots\ldots)$$

となすべきである所以を説明した。この理を相關理論に移行せしめれば、我々は一變數と二變數又はそれ以上の變數との間の相關を取扱はねばならぬことは明かである。前例について言へば、賃銀と相關するものはは單に物價のみではなく、勞務者の年齢・能率・體性とか、會社の成績或ひは規程なぞは直接の關係を持つて居り、更には勞働組合や國家の影響も少なからざる力を及ぼしてゐる。これらのうちには數字的に評量し得ない要素も少くないから、その總べてを計算に入れることは勿論不可能であるが、可能なる範圍内に於ては成るべく多くを採り入れることによつて、賃銀決定の眞相に接近することが望ましい。多元相關 (multiple Correlation) 及び部分相關 (Partial Correlation) の計算はこの要求に應ぜんとするものである。

併し乍ら、極めて多くの變數を採り入れることは、實際問題としては不可能である。我々はこの際にも單純化に訴へざるを得ないのであつて、複雜な關係といひ乍ら、實は一變數と他の二變數乃至三變數との關係に限定するのを原則とする。加之計算に於ては他の諸變數の一を可變とし他を一定と假定するとか、或ひは他の諸變數は直線的相關に在るものと假定するが如き方法を採るのである。前者を部分相關、後者を多元相關といふ。いまx, y, zの間に相關あるものとすれば

(1) 部分相關を求める式は

$$r_{12\cdot 3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{(1-r^2_{13})(1-r^2_{23})}}$$

(2) 多元相關Rを求める式は

$$1-R^2_{3(12)}=(1-r^2_{31})(1-r^2_{32\cdot 1})$$

である。$r_{12\cdot 3}$ とはzを一定とした場合にxとyとの間に成り立つ部分相關、r_{13} とはxとzとの間の相關係數、$R_{3(12)}$ とはxとyとが直線關係にあるものとして、それとzとの間に成り立つ多元相關を意味する。これらの問題は初等統計學の範圍ではないから、こゝでは一切省略する。實際にこの種の計算は餘り利用されてゐない。

四　順位差相關係數（スピアマンの方法）

上記ピアソンの方法の一變形として、項の大さの順序だけから相關度を計算する方法がある。最初から順序だけしか與へられてゐない資料は少くない。學業の席次なぞその例であるが、かやうな場合にも相關度の問題はいつも起つてくる。首席だつた人は概していつも上位にゐるだらう

し、落第するやうな人はいつも下位にゐるのが常である。點數で成績が示されてゐればピアソンの式で二度の試驗結果の間の相關度を測れるが、唯だ席次だけ發表されたときは、この方法は役に立たない。スピアマン（Spearman）はかゝる場合に利用できる方法を導いた。順位差相關法と稱せられるが、實はピアソン相關式の變形に過ぎないのである。いま甲は第一試驗では五番、第二試驗では三番だつたとすれば順位差は2である。全員について同樣に順位差が計算される。これを一般にdで示せば、順位差相關係數ρは次の式によつて求められる。

$$\rho = 1 - \frac{6\Sigma(d^2)}{n(n^2-1)}$$

ところがこの方法は、元來ピアソン式によるべき資料についても適用できる。即ち各項の大さが與へられてゐる以上、これを順位に書換へることは容易だからで、而もこれによつて計算は著しく簡單になるのである。上記の例についてこれを求めれば、次のやうな結果になる。

年次	物價	賃銀	物價の順位	賃銀の順位	d	b²
1	158	101	9.5	6	3.5	12.25
2	146	96	6	3.5	2.5	6.25
3	141	96	4	3.5	0.5	0.25
4	134	95	2	1.5	0.5	0.25
5	130	95	1	1.5	—0.5	0.25
6	137	100	3	5	—2	4
7	150	103	7	7	0	0
8	158	104	9.5	8.5	1	1
9	153	104	8	8.5	—0.5	0.25
10	145	105	5	10	5	25
						49.5

$$\rho = 1 - \Sigma \frac{6\,\Sigma(d^2)}{n(n^2-1)} = 1 - \frac{6 \times 49{,}5}{10 \times (10^2-1)}$$

$$= +0.709 \fallingdotseq +0.71$$

即ちこの場合にはピアソン式と全く同じ結果となつた。勿論總べての場合さうとは限らない。

	X	X
甲	3	8
乙	10	9
丙	11	10

例へば上の表から順位相關係數を求めれば、X系列もY系列も甲乙丙の順序は、1、2、3、となり、従つて順位差は零となるから、相關係數は一となる。然るにこの二系列は完全な比例關係ではないから、ピアソン式によれば一とはならぬ。計算すれば〇・九一八である。

順位差相關の計算で注意を要するのは順位のきめ方である。同じ大さが二つ又はそれ以上あるときは、その間に順位はないわけで、例へば2 4 4 7とあるときは、順位は1、2、3、4ではない第二と第三の項は等しいから、第二及び第三の二つの順位を平等にもつのが至當で、從つて1 2.5 2.5 4とすべきである。前掲計算例の物價及び賃銀の双方にこの手續が示されてゐる。

註　相關論の對象は必ずしも常に變數間の關係に限られてはゐない。即ち時には非數字的な屬性、例へば親子の頭髮の色の關係の如きものを問題とすることがある。學業成績を優良可等の評語で示したとき、語學のそれとの相關度を測定する場合もこの例に屬する。これを共伴屬性相關法といひ、その相關度を示す係數を共伴屬性係數又は一致係數（Coefficient of Contingency）といふ。

第十五章　Statistik より Stochastik へ

一　統計學の學的性格の變貌

馬を指して鹿といひ、羊頭を掲げて狗肉を賣るといふが、名と實の伴はないものは數へ出せばきりがない。統計學の原語 Statistik（英語の Statistics）といふ文字が、實はその適例なのである。即ち、語源的には、十七世紀の獨逸に現はれた一種の國家學に由來し、國家に顯著な事實（Staatsmerkwürdigkeiten）を記述する學問を意味した。これは別に數字を用ひたわけではなく、今日の統計學とは似てもつかぬものであつた。然るにかゝる學問はいつしか消滅した、といふよりは、寧ろ、他の諸々の學問（地理學、歷史學、經濟學、人口學等々）に分解されて了つた。偶々同じ頃英國に發生した政治算術は數字によつて集團現象の規則性を摑まんとし、こゝに統計學の實質が芽生えたのである。然るにこの政治算術といふ名稱はいつしか Statistics と置き換へられ、こゝに名實相伴はざる一つの學問が成立するに至つた。この新らしい學問は確率理論と密接に結合し、事象の規則性とその限界を確定することを以て任務とする純然たる方法の學

(Methodenlehre) である。これは國家現象のみならず、社會現象一般に、否、自然現象にさへも適用される一つの論理的形式的科學である。Statistics の名稱は全く不適當で、碩學ボルトケウィッツ (Bortkiewitz) は Stochastik と改稱すべきことを提案した。ストハスティクとは「推量する」といふ希臘文字から造られ、「確率論的基礎に立脚する大量觀察の學」といふ意味である。勿論久しく使ひ慣れた名稱は改めにくいもので依然 Statistics の文字が使はれてゐるが、內容的には既に別物であることを銘記せねばならぬ。これだけを豫備知識として、簡單に統計學の歩んだ跡を振返つてみよう。

十七世紀の中葉、獨逸では中央集權國家の成立と共に有能な政治家や行政官の養成が必要となつた。このため國內及び國外の事情を具體的に說明する一種の國勢學が成立した。最初これに體系を與へたのはヘルマン・コンリング (H. Conring, 1606—1681) であつたが、その後繼者アッヘンワル (G. Achenwall, 1719—1772) はこれに Statistik といふ名稱を與へた。國家の顯著な事實を記述する學問といふ意味である。それは主として普通の言語を用ひ、數字によつたわけではない。丁度その頃ロンドンにジョン・グラント (J. Graunt, 1620—1674) といふ有能な商人が現はれた。彼は教會に保存された出生、死亡、結婚等の記錄から統計を作製し、一見不規則なこれ

ら現象が實は驚くべき規則性を有することを發見したのである。これは明かに統計的研究で、人間の知的努力に一つの大きな途を拓いたわけである。彼の著「死亡表に關する自然的政治的觀察」（一六六二年）は斯くて統計學の最初の一里塚となつた。彼の方法は友人ペティー（W. Petty）の繼承するところとなり、同時にこの方法に對して政治算術の名稱が與へられた。ハレー慧星の發見者として有名なハレー（E. Halley）はこの方法を活用して最初の生命表を作製し、以て今日の生命保險の基礎をつくつた。

政治算術は大陸にも傳はり、遂に一七四一年、プロシャの僧ジュースミルヒ（J. P. Süssmilch）の「神序論」（詳しくは、人類の出生、死亡及び增殖から證明さるゝ、人類變化に於ける神の攝理）によつて大成されるに至つた。彼は人口統計に於ける各種の規則性を確定し、これら規則性を以て、神が地上に實現せしめんとする偉大な計畫の手段と說いたのである。例へば男女出生比が二一對二〇なること、及び男子の死亡率が幾分女子のそれよりも高いことを述べてから、彼はこれは結婚年齡に於て男女を同數ならしめ、一夫一婦制を實現せしめんとする神意の現れと解した。解釋そのものには非科學的色彩が强いが、方法そのものは極めて洗練されたもので、彼を以て統計學の祖と仰ぐ人の少くないのも尤もである。

ところが既にそれ以前から數學の世界では確率計算が考案されてゐた。それは元々賭博の當率を推定するために案出されたといふが、有名なパスカル(Pascal)、フェルマ(Fermat)及び特にヤコブ・ベルヌーキとその甥のダニエル・ベルヌーキ(Jacob u. Daniel Bernoulli)によつて著しく發展せしめられた。「觀察度數が增大するにつれて、經驗的確率は次第に先天的確率に一致する」といふ大數法則は彼等によつて證明されたのである。後にラグランジ(Lagrange)やガウス(Gauss)の如き天才的數學者は一層この理論を深化した。確率曲線はガウス曲線とも呼ばれてゐる。併し彼等の世界は純粹數學のそれであつて、確率理論が實生活にいかなる關係をもつかは全く度外視されてゐた。故にラプラース(Laplace)がその「確率に關する哲學的考察」に於て、この理論が人間行爲の判定に重要な役割を演じうることを述べたことは、異例でもあり、また劃期的意義をもつものでもある。

然るに從來の政治算術を確率論と結合し、近代的統計學の基礎をつくつた一人の偉大な頭腦が表はれた。平均人及びケトに關聯して言及したアドルフ・ケトレー(A. Quetelet)これである。彼はベルギーに生れ、最初は數學と天文學とを專攻し、パリーに遊學中にラプラース、ポアソン、ランボルト、フーリエ等と交誼を結び、數學的硏究に出精したが、彼の興味はいつしか社會

科學に移つて行つた。彼は社會的規則性を重視する餘り、社會法則と自然法則との差別を否定する態度を示し、その著「人間に就いて」は別に「社會物理學」と題されたほどである。彼の理念の中心が「平均人」にあることは既に述べた通りである。彼には多くの行過ぎがある。彼の後繼者例へばハーシェル（Herschell）やバックル（Buckle）が個人の自由意思を否定し、救ふべからざる宿命論に逸脫したのは、その影響によつてである。併し今日の統計學が彼によつて體系づけられたことは爭ふべくもない。彼は正しく近代統計學の父である。

前世紀後半に至つて、統計學は生物學者ゴルトン（F. Galton）及びピァーソン（K. Pearson）によつて飛躍的進步を遂げた。それは何れも確率論的方法であつて、ストハスティクとしての統計學の基礎はこゝに確立されたといひ得よう。最近の傾向はこれを押進めて、小さな部分から大きな全體を推定せんとする所謂小標本理論にまで發展しつゝある。

近年活躍し、または現に活躍しつゝある著名な學者は左の如き人々である。

エッヂワース（Edgeworth）、ボーレー（Bowley）、スタンプ（J. Stamp）、ユール（U. Yule）、ムーア（H. Moore）、シュルツ（H. Schultz）、パーソンズ（W. Persons）、ミルス（F. Mills）クズネ（Kuznet）、パール（E. Pearl）、ヂニ（Gini）、モルタラ（Mortara）、アフタリオン

(Aftalion)、マルシュ(March)、ウェステルゴード(Westergaard)、ボルトケウヰッツ(U. Bortkiewicz)、チュプロウ(Tschpurow)、等々。なほ計量經濟學者のこの方面への寄與は甚大で、特に指數論は彼等によつて面目を一新せんとしてゐる。極く新らしい學者については附錄の文獻について知られたい。

以上の簡單な概觀から、統計學に對する確率論の主たる使命は二つあることが判らう。一つは統計的規則性を理論的に裏づけるところの「大數法則」であり、他の一つは抽出調査の正確性を決定する「誤差理論」である。これらは高等統計學の問題であるが、問題の性質だけは、一應說明しておきたい。

二　數學的（先天的）確率

數學でいふ確率の何たるかは諸君の既に辨へて居られることであらう。錢を投げて表の出る確率は$1/2$、出ない確率は $1-\frac{1}{2}=\frac{1}{2}$、賽を投げて或る特定の目の出る確率は$1/6$、出ないそれは$5/6$である。出る確率を$p$、出ないそれを$q$とすれば

$$p+q=1 \quad \therefore \quad q=1-p$$

である。いま二枚の錢を投げれば、二枚とも表、一枚表で一枚裏、二枚とも裏三つの場合が起るが、それぞれの確率は$1/4$、$2/4$、$1/4$である。この分子の1、2、1は$(p+q)^2$の展開式即ち$p^2+2pq+q^2$の各項の係數に等しく、分母の4はこれら分子の總和である。同樣にして四枚を投げたときの五つの場合（全部表、三枚表一枚裏、二枚表二枚裏、一枚表三枚裏、全部裏）のそれぞれの確率は

$$\frac{1}{16},\ \frac{4}{16},\ \frac{6}{16},\ \frac{4}{16},\ \frac{1}{16}$$

で、前と同じく分子は$(p+q)^4$の展開式、即ち

$$p^4+4p^3q+6p^2q^2+4pq^3+q^4$$

の各項の係數に等しい。よつてn枚を投げたときの確率は$(p+q)^n$を展開することによつて求められる筈である。一般にこの式を二項式といふ。そして展開の原理は「組合せ」の理論によつて次の如く要約される。

$$(p+q)^n=p^n+np^{n-1}q+\frac{n(n-1)}{1.2}p^{n-2}q^2+\cdots\cdots+q^n$$

投錢に於ては$p=q=0.5$として、投賽に於ては$p=\frac{1}{6}$、$q=\frac{5}{6}$として計算すればよい。また、この式は一回試みたときの確率であるから、N回試みたら、$N(p+q)^n$として計算すべきである。

なほ$(p+q)^n$を展開したときの項の數は $n+1$ なることに留意されたい。

pとNとの乘積を期望値（Expectation）といふ。豫期される理論値といふ意味である。一枚の錢を十回投げたときの表の出るそれは $10\times0.5=5$回 である。四枚を百回投げたときの表三裏一の期望値は $100\times\frac{4}{16}=25$回 である。十萬枚に一本の百萬圓富籤を二枚買つたら$2\times\frac{1,000,000}{100,000}=20$圓 が期望値である。二枚百圓とすれば、みすみす八〇圓は損するものと覺悟すべきである。

確率計算には確率の加法とか減法とか、幾つかの原則があるが、こゝでは省略する。一つ注意して欲しいことは、$(p+q)^n$ の展開に於て、$p=q=0.5$ ならば完全に對稱的なること、pとqとが等しくなければ對稱的ではないが、而もnが非常に大きくなれば、それに順じて次第に對稱的になるといふことである。これはnが大なるに從つて歪度S^kが次第に小さくなるといふことで、次の如く證明される。

$$S^k=\frac{M_a-M_o}{\sigma}=\frac{np-[(n+1)p]}{\sqrt{npq}}$$

ところが分子の絶對値は1より小さいから、nが大なるに從つてS_kは零に近づく、即ち正常分布

に近づくのである。

二項分布の性質については、誤差法則に關聯して改めて述べよう。

三　經驗的確率と大數法則

右に述べた投錢または投賽に於ける確率は、別段實驗を經ないで論理的に、即ち先天的に求められるもので、これを先天的または數學的確率といふ。ところが今のような確率法則を知らない人が實際に錢を投げて見たとする。一箇を投げたとし、その結果を觀察すると、回數が少なければ結果は甚だ不規則で、何度も表ばかり出たり、裏ばかり出て、例へば十回やつて見ても、恐らく五回は表、五回は裏といふ結果は得られないであらう。ところが回數を重ねるに從つて兩者の回數は次第に等しくなることを發見するであらう。千回もやつて見れば略々表も裏も五〇〇に近くなる、即ち出現頻度は何れも $\frac{500}{1000}$ 即ち $\frac{1}{2}$ に近くなるものである。これは結局誤差が次第に相殺されるからで、僅かの回數だと表ばかり餘計に出ることは止むを得ないが、續けてゆけば次には裏ばかり出る、その次にはまた表ばかりといつた工合に、變態的出現そのものが互ひを殺し合ふのである。然らば實驗度數を無限大にすれば、結果は先天的確率と一致する筈で、いま實驗度

數をm、表の出た度數をnとすれば、次の式が成り立つ。

$$\lim_{m\to\infty}\frac{n}{m}=p$$

即ちmが大なるに從つて、比率はpに接近し、mが無限大となれば、pに收斂する、或ひは$\frac{n}{m}$の極限値はpである。十箇の錢を投げた場合も同樣で、實驗度數を無限に重ねれば、表五裏五の場合を中心とした左右對稱的な分布狀態が現はれ、その各々の組合せの出現確率は先天的確率に一致することが判るであらう。これを大數法則といふ。

して見ればわれわれは或る事象の確率は數學的にも經驗的にも求めうるわけである。その何れがより適當であらうか。惟ふに投錢や投賽の如き機械的事象は別として、殆ど大部分の事柄についてわれわれは先天的確率を知る術はないのである。例へば男女出生比を考へるに、生れる子供は男か女に決まつてゐて、恰も錢の表か裏かの問題に同じと思はれるかも知れない。併し男の生れる確率pと女の生れる確率qとが相等しいとは、どこにも定められてゐないのである。$p+q=1$即ち$q=1-p$といふことは間違ひないが、pとqの大さは先天的には言へる筈がない。男女が略々同數生れるといふことは、單に經驗から結論されるだけで、各種の生物について見れば、雌雄の割合は可成りちがつてゐる。人間の壽命、結婚年齡、火災の頻度、物價の騰落等々殆ど凡

ゆる事柄は、何れも確率的にしか言へないことは明かであらうが、同時に何れもその先天的確率は全く知られてゐないのである。

然らば實際の確率はこれを經驗的に求める外はないのであつて、先天的確率といつた抽象的理念を排し、「頻度の極限値」として確率を一般的に規定した方がよいといふことになる。ミーゼス（Mises）以來この考へ方が次第に勢力を得つゝあるのも不思議はない。

併しこの問題で最も困難な點は、經驗的確率が先天的確率に一致するためには、經驗度數を無限大としなければならぬといふことである。言ふ迄もなく、限りあるわれわれの力で現實に無限大を取扱ふことは出來ない相談である。われわれが現實に把捉できるものは必ず有限であり、而もその限界は甚だ狹いのが常である。して見れば經驗的確率も單に理念の上にしか存在しない抽象物で、その非現實的なること猶ほ數學的確率と大差はないことになる。換言すれば、大部分の事象についてわれわれは確率を云々する資格がないといふことである。だが幸ひにしてこれに對處する途は拓けてゐるのである。

その第一は、一般に觀察の正確度は觀察回數と比例するものでなく、略々その平方根に比例するといふことである。一〇〇回の實驗を一〇、〇〇〇回に增加しても、結果の正確度は一〇〇倍

されずして、$\frac{\sqrt{10,000}}{\sqrt{100}}$即ち一〇倍されるに過ぎない。して見れば觀察度數を甚だ大きくすれば、事實上は無限大にしたことゝ大差がないわけで、それから得られた結果は經驗的確率の、延いて先天的確率の、近似値と見て差支へないのである。無限大の觀察は不可能でも、甚だ多くの觀察なら不可能ではない。それは即ち大量觀察であつて、それを可能ならしめる方法が即ち統計調査なのである。われわれが敢へて統計を求める主たる理由が、事象の確率を求めるに在ることは、斯くて明かにされたと思ふ。

第二に、統計的結果の安定性を擧げることができる。統計調査が無限大の觀察に代用されるといつても、もし調査の結果が常に浮動的だとすれば、事實は代用してゐないわけである。觀察度數を無闇に增大する必要のないことは右に述べたが、ではどの程度で止めてよいか、その限界は定かでない。一年間の出生總數から男子出生の確率を求めうるかどうかは、豫め言ふことはできないのである。一年間の觀察結果は、その年の男女出生比を示すものではある。併しこの比が直ちに確率かどうかは、それだけでは判らない。もし年々の觀察の結果が相互に異るものとすれば、確率の所在は不明と言はねばならぬ。その場合には一年といふ期間を五年とか十年とかに擴大し、より大きな觀察をやつて見る外はないであらう。要するに調査の結果がいつも一致して始

めてわれわれは經驗的確率を云々しうるのであつて、この場合のみ統計的結果は卽ち一般的法則である。

いま男女出生比を見るに、年々のそれは何れの國に於ても甚だ安定的である。我國では例年女一〇〇に對し男一〇二乃至一〇三で年々の差は極めて少い。勿論人の努力によつて左右されるものは安定性が少いにきまつてゐる。乳兒死亡率の如きその適例で、これらの事柄は、それが可變的であることが人の努力を促す理由である。假りに胎兒の性を自由に決定できることになれば、男女出生比は兩親のそのときどきの希望の如何で決まり、甚だ不安定となるであらう。かような事柄には嚴密な意味に於ける確率の概念は適用できないであらう。

併し乍ら從來の經驗によれば、極めて多くの統計的結果は、たとへそれが社會的性質のものであつても、豫想外の安定性を示してゐるのである。一見規則性を有しない犯罪または自殺といふ恣意的行爲も、統計的に觀察すれば件數に於てまた內容に於て驚くべき安定性をもつてゐることが判る。社會的諸現象が自然現象に似た法則性をもつことは、特にケトレーの指摘したところで、彼がその主著を「社會物理學」(physique Sociale) と題したことが、この間の消息を端的に表明してゐる。卽ち社會現象の法則性を統計から導き出し、社會科學を自然科學の水準にまで高

ゐんとするのが彼の目的であつて、統計學は彼に於てはこの大きな使命を擔當せしめられたわけである。消費に關するエンゲルの命題、即ち所得に對する食費の割合は所得の低下と共に增大するといふことが統計的に確認される以上、これを以て確率と見ることができ、從つて社會的法則と名づけて差支へないのである。

以上の二つの理由、即ち相當に大きい觀察によれば略々經驗的確率を求めることができるといふこと、及び實際の幾多の統計的結果は著しい安定性をもつてゐるから、これを以て經驗的確率即ち確率そのものと見て差支へないといふこと、この二つの理由は正しく統計的方法の眞の理論的根據と言はねばならない。より簡單にいへば、統計學の基礎は大數法則に在るのである。

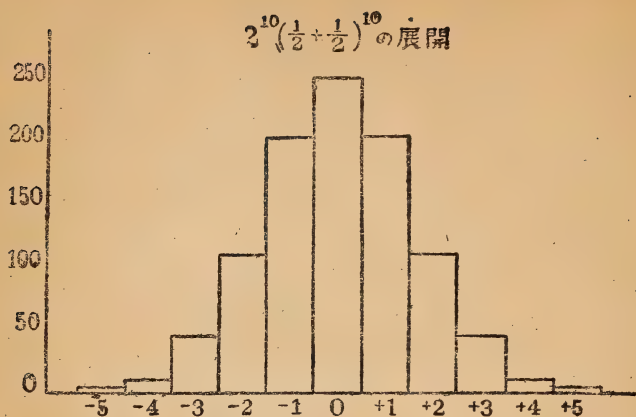
四　誤差法則と試料調査

前條によつて統計調査は可及的に大規模でなければならぬことが結論された。國勢調査その他の悉皆調査はこの線に沿つて行はれるものである。然るに既に述べた通り、かゝる調査は幾多の不便を伴つてゐる。費用勞力、迅速等々の點からは、寧ろ逆に調査範圍は可及的に縮小しなければならない。一方では調査は可及的に擴大されなければならぬといひ、他方では逆に可及的に縮

小されなければならぬといふのは、明かに矛盾であるが、この矛盾は誤差理論によつて或る程度まで克服されるのである。

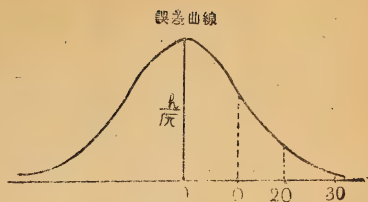
誤差とは眞實の値または理論的な値と、現實の値との開きを意味する。平均値と各項との開きを偏差と名づけたが、もしこの平均値が理論的に正しい値の場合には、偏差は即ち誤差である。誤差は英語で Error といふが、Errorには、別に誤謬、失策、過誤といつた意味もある。誤差も若しそれが觀察者の不注意、不熟練または用具の不備等から起つた場合は、寧ろ誤謬である。それは人の努力によつて限りなく排除される。ところが例へば投錢に於て一〇〇回やつて見て表が六〇回出たとすれば、一〇回は誤差であるが、これは人の罪ではない。それは人力を超えた謂はゞ秘密的な理由からそうなるので別の言葉でいへば、偶然の結果である。故にかような誤差は一般に偶然誤差と名づけられる。誤謬は一方に偏する傾きがある。物指が縮んでゐれば、結果は總べて過大となり、澤山測れば測るほどその程度は甚だしくなる。然るに誤差は偶然の結果であるから一方に偏することはない。換言すれば正負の誤差は相等しくなる傾きがある。更にまた誤差はその大きなものほど起り難い。これらのことは上記の投錢の例から容易に導き出せるのである。

十枚投錢に於て、最も期待される結果は表五裏五の場合であるから、これを標準として測れ

ば、表六裏四の場合 +1 の、表四裏六の場合は −1 の誤差を伴ふものと考へることができる。

第二十六圖

以下かようにして零を中心として −5 から +5 までの誤差ができる。その分布は第二十六圖の通りであるが、その形は零を中心とした左右對稱的な山型である。十枚投錢に於ては$\left(\frac{1}{2}+\frac{1}{2}\right)^{10}$を展開して得られる十一箇の項(−5から+5まで)を得たが、(第一圖)枚數を增大すれば項數も亦增加し、枚數を無限大とすれば、卽ち

$$\lim_{n\to\infty}\left(\frac{1}{2}+\frac{1}{2}\right)^{n}$$

とすれば、項數も無限大となつて、度數棒圖表(ヒストグラム)は完全な曲線圖表となる。(第二十七圖)これが誤差出現の一般形で、こ 曲線を誤差曲線、確率曲線、正常曲線、正規曲線、またはガウス曲線、ケトレー曲線なぞと名づける。名

稱はかく多いが何れも同じものである。その方程式は次の如くである。

$$y=\frac{h}{\sqrt{\pi}}e^{-h^2x^2}=\frac{1}{\sqrt{2\pi pq^n}}e^{-\frac{x}{2pq^n}}=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$$

第二十七圖

十枚投錢に於て誤差零なる場合の確率は $\frac{252}{1024}$、同じく誤差 +1 及び +2 の場合のそれは、それぞれ $\frac{210}{1024}$, $\frac{120}{1024}$ であるから、零から +2 までの誤差の起る確率は、これらを合計したもの、即ちその間の棒の面積を加へたもの即ち

$$\frac{252}{1024}+\frac{210}{1024}+\frac{120}{1024}=\frac{582}{1024}$$

である。正常曲線に於ては零から a までの誤差の起る確率は、定積分の方式に從つて次の式で示される。

$$\int_0^a f(x)dx=\int_0^a\frac{1}{\sqrt{2\pi pqn}}e^{-\frac{x^2}{2pqn}}$$

いまｘ座標をσを單位とし零からｘ軸に沿つて右方にσに等しい距離をとれば、その間に挾まれ

る面積は

$$\int_{0}^{a} f(x)dx = 34.134\%$$

である。同様にして 2σ をとれば 47.379%、3σ をとれば 49.865% となる。左方についても同様であるから、従つて左右について測れば

$$\int_{-\sigma}^{+\sigma} f(x)dx = 2\times34.134\% = 68.268\%$$

$$\int_{-2\sigma}^{+2\sigma} f(x)dx = 2\times47.379\% = 94.758\%$$

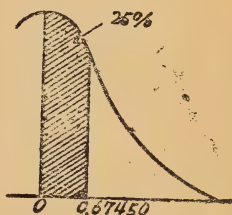
$$\int_{-3\sigma}^{+3\sigma} f(x)dx = 2\times49.865\% = 99.730\%$$

となる。換言すればこの曲線の包む全面積（即ち起りうる全誤差は殆ど -3σ から $+3\sigma$ までの間に限られるのである。）また零から左右に σ の〇・六七四五倍をとれば、その間に含まれる面積は全體の丁度半分となる。即ち

或ひは片方だけについて計算すれば（第二十八圖）

$$\int_{-0.6745}^{+0.6745} f(x)dx=\frac{1}{2}$$

$$\int_{0}^{+0.6745} f(x)dx=\frac{1}{4}$$

第二十八圖

これは全體の半分はこの範圍内の誤差をもち、從つて他の半分はそれ以上の誤差をもつといふことで、換言すれば全體の中から任意にとつた一つがもつ誤差は、この範圍内であるか又はないか、その確率は等しいといふことである。故にこの

$$0.6745\sigma$$

を特に**確率誤差**といふ。

起りうる誤差の大さは略々3σに等しいといふ上記の命題は、統計學に於て幾多の用途をもつてゐる。與へられた度數系列が略々正常分布をなすときは、即ち分布が正常曲線に近い形を示すときは、その系列の平均値mにこの3σを添附し

$$m\pm 3\sigma$$

として現はしておけば、分散の程度は極めて明瞭となるのである。例へば一〇〇人の身長（身長の如きは常に略々正常分布を示す）の平均が一五〇糎、標準誤差が五糎とすれば、この全員の身長は150±3×5即ち一三五糎から一六五糎までに限られ、一三五糎よりも小さな人や一六五糎よりも大きい人は殆どないといふことを意味する。勿論これは系列が正常分布に近い場合に限られるから、甚だしく歪んだ分布に於ては適用されない。例へば賃銀の平均値が二〇〇〇圓、σが一五〇圓としても、全員の賃銀が2000±3×150即ち一五五〇圓から二四五〇圓までだとは斷定できない。即ち資料の性質によつて適用のできるときとできないときがあるが、與へられた系列の分布狀態がはつきりしてゐないときには、正常分布を假定するのが常であるから、また事實正常分布にちかいものが極めて多いから、利用の範圍は甚だ廣いのである。先に述べた回歸線から相關係數を導く場合など、正にこの命題の典型的應用例といへる。その應用ついては標本調査を說明するときに述べることとしたい。

確率誤差について言へば、上例に於て平均身長に確率誤差を添附した150±0.6745×5即ち一四六・六糎は、全員の半數の身長の範圍を示すものである。換言すれば、全員中から任意にとつた一人の身長は、この範圍内なる場合と範圍外なる場合と相等しいといふことである。範圍外と

は勿論 $3\sigma-0.6745\sigma$ である。

正常曲線によつて示される正常分布は、二項式 $(p+q)^n$ に於て p と q とが等しい場合である。よつて二項分布についての一般的性質は當然正常分布にも當嵌まる。最も重要な性質は次の二つであらう。

(1)二項分布の算術平均値 m は np である。故に正常分布では p は $\frac{1}{2}$ であるから、算術平均は $\frac{n}{2}$ である。

證明。$m=\frac{1}{N}(p+q)^n=np[q^{n-1}+(n-1)q^{n-2}p+\cdots+(n-1)qp^{(n-1)-1}+\cdots]$

$=np(q+p)^{n-1}=np$

例へば投賽に於ては $p=\frac{1}{6}$ であるから、n回の試行で或る目の出る度數は平均 $n\times\frac{1}{6}$ である。ところが投錢に於ては $p=\frac{1}{2}$ であから、表の出る平均度數は $\frac{n}{2}$ となる。千枚づゝ投げれば表は毎回平均五百枚出る。

(2)二項分布の標準偏差σは $\sqrt{npq}$ である。故に正常分布のそれは $\sqrt{n\frac{1}{2}\cdot\frac{1}{2}}=\frac{\sqrt{n}}{2}$ である。

證明。$\sigma^2=\frac{1}{N}\Sigma f(x-A)^2-(m-A)^2$ であるから、$A=0, m=np$ として變形すれば

$\sigma^2=\frac{1}{N}[0^2\times q^n+1\times nq^{n-1}p+\cdots+(n-1)^2nqp^{n-1}+n^2p^n]-(np)^2$

$$=np[q^{n-1}+\cdots\cdots+(n-1)^2qp^{n-2}+np^{n-1}]-(np)^2$$

$$=np[(q+p)^{n-1}+(n-1)p(q+p)^{n-2}]-(np)^2$$

$$=np[1+(n-1)p]-(np)^2=np-np^2=np(1-p)=npq$$

故に正常分布では　$\sigma^2=n\cdot\frac{1}{2}\cdot\frac{1}{2}$或ひは　$\sigma=\frac{\sqrt{n}}{2}$である。例へば十枚の投錢に於ける標準偏差は$\frac{\sqrt{10}}{2}$である。正常分布の標準偏差を特に標準誤差といふ。回歸線から相關係數を誘導する場合に推算の標準誤差Sを用ひたのは、正常分布を假定したからである。即ちXの或る値に對應するYの値は、回歸線を中心として上下に正常分布をなすものと假定したわけである。

以上が誤差法則の大樣である。これが部分調査とどんな關係をもつであらうか。これは所謂數理統計學の最重要な問題で、到底本書の如き入門書では取扱へないが、比較的簡單で而も利用性の多い一例を擧げて見よう。

平均の試料誤差

極めて大なる集團について算出した算術平均値をm、その標準偏差をσとする。いまこの集團をa個に等分し、その各々に含まれる單位數をnとすれば、「觀察の正確度は觀察單位數に比例する」といふ原則によつて、これらa個の平均値の標準偏差は

$$\frac{\sigma}{\sqrt{n}}$$

である。故にaの數を減らし、從つてnを大きくすれば、それだけ標準偏差は小さくなる。何れにせよ全體の平均と僅かn人から得た平均とは

$$m \pm 3\frac{\sigma}{\sqrt{n}}$$

以上の誤差は先づ無いと見て間違ひないのである。もしこれ以上の差があつたら、部分のとり方が不公平だつたことを意味する。即ち部分から計算した平均値が能く全體を代表してゐるかどうかゞ判るのである。

併し多くの場合、全體（母集團）の平均値や標準偏差は判つてゐないのが原則である。ゐないからこれを部分調査から知らうとするのであつて、判つてゐる位ゐなら、部分調査など不用な筈である。そこで實際問題としては部分から求めた標準偏差σ'を以て之に代用するのである。nが比較的大きければσ'との差は無視できる。よつていま四百名の成年男子について身長を調査し $m'=160\text{cm}\ \sigma'=5.5\text{cm}$ を得たとすれば、母集團の平均との差は

$$m \pm 3 \times \frac{\sigma'}{\sqrt{400}} = 160 \pm 3 \times 2.75$$

即ち八・二五糎を出でることは先づないと見てよいわけである。また確率誤差を附して

$$m' \pm 0.6745 \times \frac{\sigma'}{\sqrt{400}}$$

とすれば、この範圍内に在る確率は0.5なることを意味する。これらの値は何れもσ'さへ判れば直ちに計算できるから、mを發表するに當つては一般にm'にσ'を附して

$$m' \pm \sigma'、\text{即ちこの例では} \ 160\text{cm} \pm 5.5\text{cm} \ \text{とする。}$$

全體の平均値と部分のそれとは、部分と全體が等質ならば、右の如く$3\frac{\sigma'}{\sqrt{n}}$以上の差はないものと見てよい。故にもしそれ以上の差が起つたら、部分は全體と質を異にすると判定できよう。異質といふ意味は、單なる偶然からはこれだけの差は起り得ないから、何等か本質的な相違があるにちがいないといふことである。x歳の全國の學童平均身長と某郡のそれとにもし斯かる差があることが發見されたら、その郡には何等か特殊の事情があるものと推定されよう。

右では全體の平均と部分のそれとを問題としたが、次に全體から二つの部分をとつて、その各々について求めた平均値を考へて見よう。二つの部分の大さをそれぞれm及びn、標準偏差をそれぞれσ'及びσ''とすれば、それぞれの標準誤差は上記の如く

$$\frac{\sigma'}{\sqrt{m}},\quad \frac{\sigma''}{\sqrt{n}}$$

である。ところが二つの平均値の差の標準誤差は$\sqrt{\frac{\sigma'^2}{m}+\frac{\sigma''^2}{n}}$である。故に二つの平均値を

a′及びa″とすれば、その差をこれと比較すればよい。即ち

$$\frac{a'-a''}{\sqrt{\frac{\sigma'^2}{m}+\frac{\sigma''^2}{n}}}$$

を計算し、これが 3σ より小さければ、兩者の差は偶然と見てよく、それより大きければ、この二つの部分は本質的に何等か、異るものと推定せねばならぬ。

これらは試料理論の一端に過ぎない。統計調査が大規模な悉皆調査から極く小さな試料調査に移りつゝあることは既に述べた。試料理論は實に今日の統計學の中心課題である。アメリカではこの研究は戰時中に著しく發達し、先般統計使節デミング博士の來朝によつてその眞相は可成り詳細に判つてきた。我國の統計科學研究會は特にこの問題に全力を集中しつゝある。今後の發展は期して待つべきものがあると信ぜられるが、私は諸君がこのさゝやかな入門書を一里塚として、壯大な統計學の殿堂に進まれんことを希望して歇まない。

附論　近似値とその計算

一　正確値の計算

統計的研究に於ては數字の計算は中心的課題である。數字の運算は國民學校の教程となつてゐるほどで、至極簡單な筈だが、これが思ひの外簡單でないやうである。乘除を先にし加減を後にするといふ算術第一課で既に間違ひを犯す人すら少くないところを見ると、數字の計算に關して一應の注意を喚起して置く事は必ずしも徒爾ではあるまい。

加算と減算は珠算を、乘算と除算は計算尺又は、計算器を用ふるのが最も便利であるが、その他各種の計算表を具へて置けば大いに手數を省く事が出來る。計算表には金利や年金等に關するものや各種統計々算に關する特殊の表など種々あるが、一般の數字計算表としては整數計算表と對數表とがある。前者は整數の平方、立方、平方根、逆數、階乘を纏めたもので、例へば $\sqrt[3]{3354}$ を求めるのに實際にこれを運算したら大變な手數であるが、整數計算表を一覽すれば直ちにそれが 14.968824 なる事が判る。後者即ち對數表は、常用對數、三角函數の對數及び眞數の如きを表と

したもので、これなくして複雜な乘除計算は不可能にちかい。併し平方、立方、平方根または立方根の程度なら整數表を利用した方が早い。また割算は逆數の掛算とした方が樂で、このためには上記の整數表の逆數の項か、特別の逆數表例へば Cotsworths Reciprocals を利用すべきである。普通の整數表としては次の二つがよい。

東京工學研究會編「整數の計算表」

Barrow's Tables of Squares, Cubes, Square Roots, Reciprocels, London, 1930

本書の冒頭で私は我々の日常出會ふ數字は大部分は正確な數字ではなくて近似値に過ぎないことを指摘した。計量は或る程度以上には正確に行ふことが物理的に不可能であり、計數は、小數については別としても、大數については手續上いろいろの手落ちが起り易いからである。併し何れの場合にも、注意さへすれば益々正確に近づきうることは確かで、精密檢査では何百分の一度の温度や何千分の一粍の厚さなぞが測られてゐる。圓周率πは圓周の長さを直徑の長さで割つたもので、この割算はいつ迄續けて行つても割切れない。そして續けて行けば行くほど正確になるのであつて、今日では小數點以下七百七桁までの値が判つてゐるさうである。科學的正確を尙ぶ場合には、斯くの如き極端な近似値も必要であらうが、併し日常の大部分の目的に對しては、單

に必要でないのみか、寧ろ迷惑だとも言へるのである。いゝ加減なところで妥協して大雜把な近似値で濟ませるのが、結局は生活を順調に進行させる要諦なのであつて、大晦日に生れた赤坊が翌日は二歳になつても人は不思議としないのである。尤も一寸の蟲にも五分の魂とか白髮三千丈の類ひとなれば、素より數字の意味はある筈もない。こゝでいふ近似値とは數學的のそれを指すこと勿論である。

近似値については二つの問題が起る。一つは、近似値そのものの加減乘除はどうするかで、その說明は後廻しとし、もう一つは、欲すればどこ迄も正確に近づく計算を敢へて或る限度で打切つて實用的近似値を求めることである。所謂四捨五入これであるが、併し幾桁目についてこれを行ふべきかの原則はない。πを四捨して3.14とするか、五入して3.1415とするか、或ひはより下位まで計算するかは、問題によつて決める外はない。併しこれに關聯して私は百分比%の計算について一言注意したい。

百分比は比例數の一種で、全體をA、或る事柄をBとしたとき

$$\frac{B}{A}\times 100$$

の形で示されるものをいふ。AもBも共に正確な數とすれば、細く計算すればするほどよいわけ

で、小數以下一桁よりも二桁がよく、二桁よりも三桁がよいに決つてゐる。數學的には正にさうであるが、併し實際には必ずしもさうではない。これは百分比なるものの本質を一考すれば判ると思ふ。言ふ迄もなく百分比は複雜な比率を見易い形に改めたもので、見易いためには微細點は省略されるのが當然である。省略がいやなら始めから比率なぞに換算しないがよい。即ち小數以下幾桁も並べた百分比は本來の目的と背馳するわけで、相當詳しく比率を示したいなら、千分比、萬分比、乃至十萬分比を使へばよい。百分比は小數以下幾桁で止めねばならぬといふ約束はないが私はせい〴〵一桁で打切るべきだと考へてゐる。殊に全體の數が極めて少い場合には成るべく大雜把な數字の方がよいわけである。極端な場合には全體が一〇〇に足りないものを%に改めることがあるが、斯かる方法は寧ろ百分比の濫用と言へよう。一五人のうち四人近眼があれば、確かに二六、六%强であるが、この數字が、果して百分比の目的に應ふであらうか。百分比は元々大數的事象の集約的表現として使用すべきもので、この例の如く、微少な數字を强いて擴大して見せる百分比は針小棒大以外の何物でもないのである。

說明が岐路へそれたが、本筋へ戻らう。複雜な計算に於ける近似値は高等數學によらねばならぬから、專門書に讓るとして、例へば $\sqrt[5]{0.999}$ の近似値計算を前に出た二項定理によつて行つ

て見よう。二項定理とは既に說明した通り、

$$(a+b)^n=a^n+na^{n-1}b+\frac{n(n-1)}{1.2}a^{n-1}b^2+\frac{n(n-1)(n-2)}{1.2.3}a^{n-3}b^3+\cdots\cdots$$

であるから、0.999=1−0.0001 と書換へれば

$$\sqrt[5]{0.999}=(1-0.001)^{1/5}=\left(1-\frac{1}{1000}\right)^{1/5}$$

即ち前式を $a=1,b=-\frac{1}{1000}$, $n=\frac{1}{5}$ としたものに等しい。故に求める答は

$$1-\frac{1}{5}\ \frac{1}{1000}+\frac{1/5(1/5-1)}{1\cdot2}\left(\frac{1}{1000}\right)^2-\frac{1/5(1/5-1)(1/5-2)}{1.2.3}\left(\frac{1}{1000}\right)^3+\cdots\cdots$$

$$=1-\frac{1}{5000}-\frac{2}{2500000}-\frac{6}{125000000000}-\cdots\cdots$$

となる。大體の近似値でよいなら例へば最初の二項をとつて 0.9998 とする。この答は第三項までとつたものに比して二千五百萬の一だけ多いが、この程度の誤差なら大抵の場合無視してよいであらう。同様にして 12^n は $10^n(1+0.2)^n$ として、$\sqrt{12}$ は $\sqrt{9+3}=3\times\sqrt{1+1/3}=3\times(1+1/3)^{1/2}$ として容易に計算できる

二　近似値の計算

近似値に關する第二の問題は、近似値そのものゝ運算はどうするかといふことである。前述し

た四捨五入は、運算を根氣よく續ければ續けるほど正確になるものを、唯だ便宜上中途で打切るに過ぎないから、どこで打切るべきはかその場その場の必要に應じて決める外はなく、一般的な規則といふが如きものはあり得ない。然るに近似値そのものを運算する場合には、事情は全く違ふのである。いま私が幹の太さを測つて六八糎の數字を得た。之を3.1416で割つて直徑を21.64502とすれば、一見甚だ正確のようだが、實は當らない。第一に3.1416は圓周率の近似値であり、第二に六八糎といふ太さも亦近似値に、而も恐らくは可成り怪しい近似値に決つてゐる。不正確なもので不正確なものを割る以上、或る程度以上の計算は全く無意味であつて、無駄といふよりは寧ろ間違であり、文字通りの蛇足である。成績點の合計を課目數で割つた平均點なるものも、もし81.236なぞとすれば、矢張り同樣である。この場合には分母たる課目數は正確な數であるが成績點そのものは我々教師から言ふのは憚りがあるが實は可成り曖昧なのである。

この間の消息を數字的に說明すれば次の通りである。即ちいま2が近似値だとし、これを3で割る場合に0.666666……としたところで、元々分子たる2が正確な數ではない以上、この答は一見正確でも、實際には少しも正確ではないのである。蓋し、2が近似値だといふ意味は、眞の値は$2\pm\Delta x$といふことであるから、もしこの2は實は2.01だとすれば答は0.67となるべ

く、又もし2は實は 1.99 だとすれば答は 0.663333……となるべく、更にもしこの2は四捨五入された近似値だとすれば、その眞の値は 1.5 から 2.499……までの間の或る値であつて、從つて3で割つた値は 0.5 乃至 0.833……の間の或る數である。故に $\frac{2}{3}$ を克明に計算して 0.66666……とすれば寧ろ間違ひである。こういふのを恐らく馬鹿正直といふのであらう。然らば近似數の計算はどこで打切るべきか。その原則は次の四つから成る。

（A） 乘算

$a \times b$ に於て a 又は b が近似數でそれが n 箇の有效數字から成るとすれば、答は n 箇の有效數字を持たねばならぬ。

この命題を說明するには豫め有效數字の何たるかを知らねばならぬ。小數以下の位どりを決定するための零なる數字（例へば 0.005 に於ける 0.00）を除き、一切の數字を有效數字 (Significant figures) といふ。故に 352 は三つの有效數字、0.023 は二つの有效數字から成る數である。205 の有效數字は三つ、0.20 のそれは二つである（0.20 の最初の0は單なる位どりのためであるから有效數字ではないが、最後の0は右の數が 0.21 でも 0.22 でもないことを示す數字で、從つて有效數字である）。

そこで最初の問題に立ち戻つて 167×4.352 を求めて見よう。いきなり計算すれば 726.784 となるが、いま 167 が近似値だとすれば、その有效數字は三つであるから、答は 727 とせねばならぬ。何故かといへば、例へば 167 が近似値だといふ事は、その最後の數字即ち7が四捨五入されたものである事を意味するから、從つて右の 167 は 167.499……と 166.5 との間の或る數である。そこで 167×4.352 は 167.499……×4.352 即ち 728.955……と、166.5×4.352 即ち 724.608 との間の或る數でなければならぬ。見易く書き列べれば

$$167\times4.352=\begin{cases}728.955\cdots\cdots\\724.608\end{cases}$$

となつて、信用しうるのは最初の二桁のみであり、三桁目は既に4なのか8なのか、或ひはその間なのか全く不明である。よつて 16.7×4.352 を一度正直に計算して得た 726.784 の三桁目以下は四捨五入するのが最も安當な方法となる筈で、これを行へば 727 となるのである。即ち167が近似數たる限りは、その有效數字と等しい三桁の 727 が最も合理的な答である。

故にもし反對に 4.352 が近似値ならば 167×4.352 は 726.8 とさるべきである。同樣にして例へば三百五十六圓の 16% を求めるのに、もしこの 16% が近似値であれば、答は五十六圓九十

六錢ではなくて、それを有效數字二つに四捨五入した五十七圓でなければならぬ。蓋し 16% は二つの有效數字しか持たないからである。

もし $a \times b$ に於て a も b も共に近似値だとすれば、有效數字の少い方の數が持つ有效數字の數だけ取ればよい。例へば 36×42.5 に於ては一方は二つ、他方は三つの有效數字を持つから、答は 1530 ではなくて 1500 でなければならない。

(B) 除算

乘算と同樣、除算に於ても有效數字の少い方の數によつて答を決定せねばならぬ。例へば $3754 \div 18$ はもし兩者とも正確な數字であれば、答は 209.5556 であるが、もし 3754 は正確で 18 は近似値とすれば 210 たるべく、その逆の場合は 208.6 でなければならない。

(C) 平方根

n 箇の有效數字から成る數の平方根は同數の有效數字とされねばならぬ、即ち $\sqrt{435}$ はこれが正確な數字であれば 20.856654 であるが、それが近似値であれば 20.9 とせねばならぬ。

(D) 加算及び減算

253.3 と 25.2562 が共に正確値であれば和は $253.3+25.2562=278.5562$ であるが、いま兩

者又は一方が近似値であれば、278.6 とされねばならぬ。同様に 12%＋13.2% は兩者又は一方が近似値ならば 25% なるべく、25.2% ではない。これは要するに赤坊が一人生れたからといつて、四百萬の東京都人口が四百萬一になつたとはいへないのと同じ理由に基くのである。減算の場合にも全く同様である。

參考文献（比較的最近のもの）

一般的なるもの

森田優三　統計學汎論（新經濟學全集）　日本評論社
杉・榮　理論統計學研究　昭15　立命館
小林新　統計學大綱　昭14　泰文社
宗藤圭三　統計學通論　昭14　有斐閣
蜷川虎三　統計學概論（岩波全書）　昭9　岩波
寺尾琢磨　統計學要論　昭22　慶應出版社
近藤忠雄　計數の統計學　昭19　岩波

米澤治文　統計の諸問題　昭22　有恒社

有澤廣巳　統計學要論（上）昭21　明善社

R. v. Mises, Probability, Statistics and Truth, London, 1939.

Davis-Nelson, Elements of Statistics, with Application to Economic Data, Bloomington, 1935.

A. E. Treloar, Elements of Statistical Reasoning, 1939.

Peters-Voothis, Statistical Procedures and their Mathematicl Bases, 1940.

F. C. Mills, Statistical Methods applied to Ecomics and Business, London, Rev edit. 1938,

R. A. Fisher, Statiscal Meethods for Research Workers, Edingburgh, 1936.

H. M. Walker, Mathematics, Essential for Elementary Statistics, Newyork, 1934.

E. Wagemann, Die Zahl als Detektiv.; Heitere Plaaderei

über gewichtige Dinge, Hamburg, 1938.

——, Narrenspiegel der Statistik' Die Umrisse eines statistischen Weltbildes, 1935.

Beiträge zur Deutschen Statistik; Fetsgabe für F. Zizek, Leipzig, 1936.

W. Winkler, Grundriss der Statistik, Erster Teil, Theoretische Statistik, Berlin, 1931.

O. Donner, Statistik, 1937.

Czuber-Burkharat, Die statistischen Forschungsmethoden, Wien, 1938.

Westergaard u. Nybolle, Grundzüge der Theorie der Statistik, jena, 1928.

L. D. Bernonville, Initiation à l'Analyse Statistique, Paris, 1939.

L. March, Les Principes de la Méthode Statistique, Paris, 1930.

統計數理論に關するもの

渡邊義勝　最小自乘法及統計　昭18　丸善

佐藤良一郎　數理統計學　昭22再版　培風館

安川・米田　統計數理　昭19　同文書院

小倉金之助　統計的研究法

伏見康治　確率論及び統計論　昭17　河出書店

岡谷辰治　計算法、確率、統計

中川友長　統計研究法の基礎　昭3　日本評論社

J. F. Kenney, Mathematics of Statistics, Part I & II. Illinois, 1939.

O. N. Anderson, Einführung in die mathematische Statisfik, Wien, 1935.

H. L. Lietz, Handbook of Mathematical Statistics, Boston, 1924.

A. C. Aitken, Statistical Mathematics, Edinburgh, 1939

E. Czuber, Wahrscheinlichkeitsrechtnung und ihre Anwendung auf Fehlerausgleichung. Statistik und Lebensversicherung, 2 Bde. Leipzig, 1938.

C. V. L. Charlier, Vorlesungen über die grundzüge der mathematischen Statistik, Lund. 1920.

G. Darmois, Statistique mathematique, Paris, 1928.

C. Jordan, Statistique Mathematique, Paris, 1927.

統計學史に關するもの

H. M. Walker, Studies in the History of Statistical Method. Baltimore, 1929.

H. Westergaard, Contributions to the History of Statistics, 1932. (森谷喜一郎譯、統計學史、昭18、栗田書店)

調査及び統計表に關するもの

F. Zizek, Wie statistische Zahlen entstehen; Die entscheidenden methodischen Vorgänge, Leipzig, 1937.

S. Koller, Graphische Tafeln zur Beurfeilung statistischer Zahlen. Dresden, 1940.

Walker & Durost Statistical Tables; their structure and Use, 1933.

定期刊行物

日本統計學會年報　Journal of the Royal statistical Society.

Journal of the Ameircan Statistical Association.

Annals of Mathematical Statistics.

昭和二十四年九月一日印刷
昭和二十四年九月十日發行

統計學入門
定價 百貳拾圓

著者 寺尾琢磨

發行者 岩瀬利吉
東京都千代田區神田駿河台三の七

印刷者 小野總次
東京都目黑區上目黑三の一九〇八

發行所
東京都千代田區神田駿河台三の七
株式會社 廣文社
電話神田(25)三〇二二番
振替東京一一四八六八番
會員番號A一一一〇二七番